INDEPENDENT RESEARCH
and
INDEPENDENT
EXPLORATORY DEVELOPMENT
—— FY91 ANNUAL REPORT ——

AD-A248 331
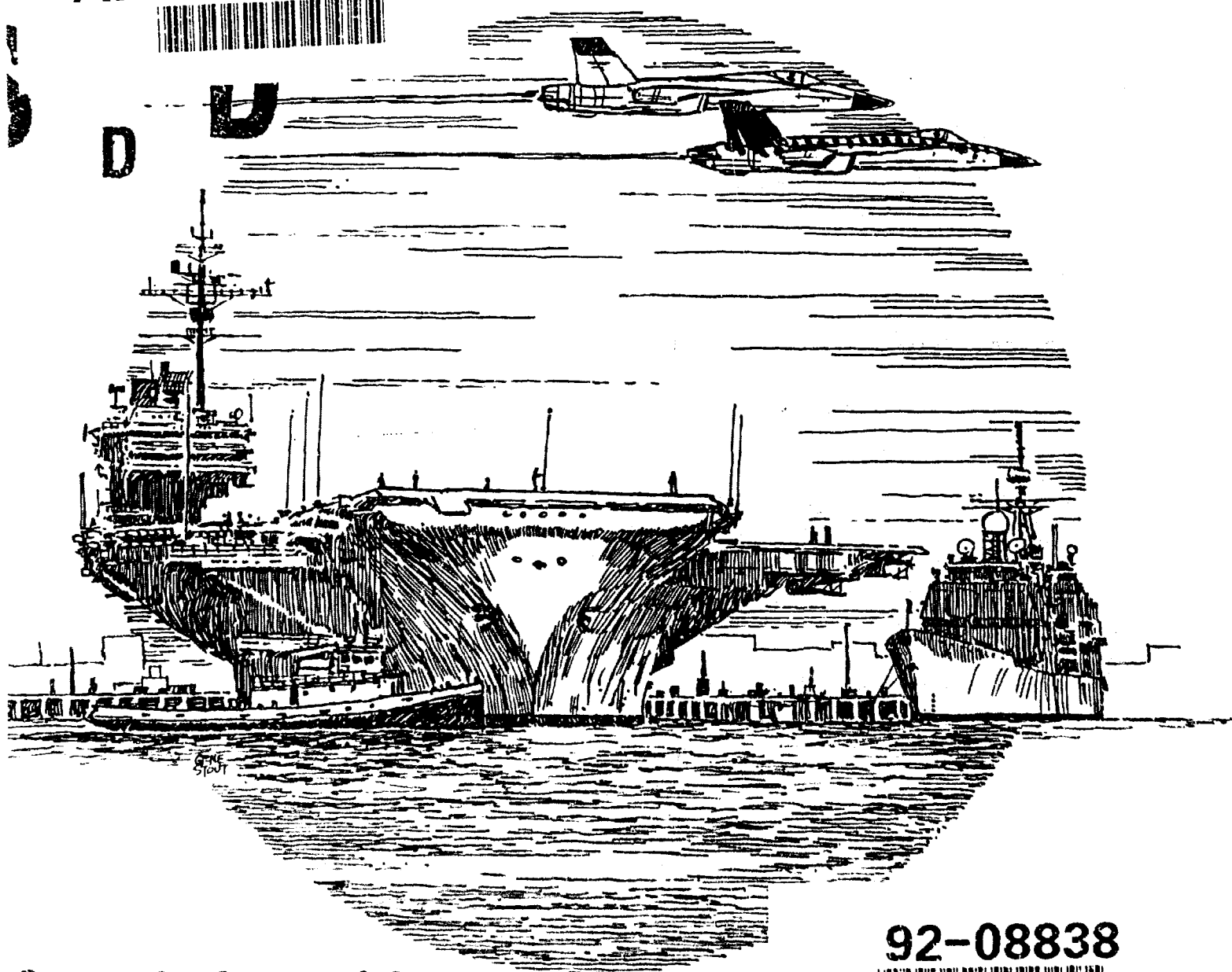
92-08838

Independent Research and Independent Exploratory Development Programs: FY91 Annual Report

William E. Montague
(Editor)

Reviewed by
Richard C. Sorenson
Technical Director (Acting)

Accesion For		
NTIS CRA&I	☑	
DTIC TAB	☐	
Unannounced	☐	
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

Approved and released by
Thomas F. Finley.
Captain, U.S. Navy
Commanding Officer

Approved for public release; distribution is unlimited.

Navy Personnel Research and Development Center
San Diego, California 92152-6800

Contents

Independent Research and Independent Exploratory Development Programs: FY91 Annual Report

Introduction

The Technical Director encourages scientists and engineers at the Navy Personnel Research and Development Center (NPRDC) to generate new and innovative proposals to promote scientific and technological growth in the organization and the development of knowledge and technology of interest to the Navy. Support for this is provided by discretionary funding furnished by the Independent Research (IR) and Independent Exploratory Development (IED) programs of the Office of Naval Research and the Office of Naval Technology. These programs support initial research and development of interest to the Navy with emphasis on the NPRDC mission areas of the acquisition, training, and effective utilization of personnel.

Funds are provided to the Technical Directors of Navy Laboratories to support innovative and promising research and development outside the procedures required under normal funding authorization. The funds are to encourage creative efforts important to mission accomplishment. They enable promising researchers to spend a portion of their time on examining the feasibility of self-generated new ideas and scientific advances. They can provide an important and rapid test of promising new technology and can help fill gaps in the research and development program. This may involve preliminary work on speculative solutions too risky to be funded from existing programs.

The funds also serve as means to maintain and increase the necessary technology base skill levels and build in-house expertise in areas likely to become important in the future. These programs contribute to the scientific base for future improvements in the manpower, personnel, and training system technology and provide coupling to university and industrial research communities.

The FY91 IR/IED programs began with a call for proposals in June 1990. Technical reviews were provided by supervisors and scientific consultants and six IR and four IED projects were funded. An invitation to prepare a special issue of an international journal, during FY91, provided an opportunity to examine the effectiveness of a technique to help in technology transfer. This report documents the results and accomplishments

of these projects. Dr. William E. Montague administers the IR and IED programs, coordinating project selection, reporting, and reviewing to assure an innovative and productive program of science and technology.

Tables 1 and 2 list the projects active during FY91 and those supported in FY92. Two papers, one IR and one IED, chosen by the Technical Director as "Best Papers of 1991" are presented. Subsequent pages, which were written by the principal investigators of each project, contain brief reports of research progress during FY91.

Table 1

Independent Research
Work Units for FYs 91 and 92
(PE 0601152N)

Work Unit	Title	Principal Investigator	Internal Code	Telephone (619) 55 or DSN 55	FY Funding (K) 91	FY Funding (K) 92
0601152N.R0001.01[a]	Brain mechanisms and cognition: Wavelets	Trejo	13	37711	0.0	42.2
0601152N.R0001.02	Cognitive resources and learning strategies	Tatum Nebeker	16	37955 37979	0.0	48.0
0601152N.R0001.04	Information processing and decision style	Morrison	12	39256	0.0	30.0
0601152N.K0001.09	Feedback in computer-based training	Cowen	13	37698	56.0	2.0
0601152N.R0001.10	Neural networks as an alternative to regression	Wilkins Dickieson	13	37618 39270	50.0	55.0
0601152N.R0001.11	Neural network modeling of skill acquisition	Gollub Dickieson	13	37849 39270	60.0	13.0
0601152N.R0001.13[b]	Prediction models for dichotomous criteria	Sands Wilkins	12	39266 37618	87.0	0.0
0601152N.R0001.14	Neural network analysis of physiological data	Ryan-Jones Lewis	13	37710 37711	55.0	0.0
0601152N.R0001.15	Experienced-based career development-III	Morrison	12	39256	25.0	0.0
0601152N.R0001.16	Decision making	VanMatre	13	39272	16.0	0.0

[a]Additional matching funds obtained from the Office of Naval Research.
[b]Transitioned to 6.2 project Exploiting Manpower Data.

Table 2

Independent Exploratory Development
Work Units for FYs 91 and 92
(PE 0602936N)

Work Unit	Title	Principal Investigator	Internal Code	Telephone (619) 55 or DSN 55	91	92
0602936N.RV36I27.01	Linking biodata and personality	Booth-Kewley	01E	39251	0.0	27.0
0602936N.RV36I27.03	Enlisted requirements model	Krass Thompson	11	37895 37925	0.0	50.0
0602936N.RV36I27.10	Decomposition methods	Thompson Krass	11	37925 37895	49.0	0.0
0602936N.RV36I27.12	Effects of administration method and anonymity/ identification on survey responses	Booth-Kewley	01E	39251	69.0	0.0
0602936N.RV36I27.13	Cognitive and motivational effects of employee involvement interventions	Sheposh	16	37947	73.0	29.0
0602936N.RV36I27.14	Transitioning research	Montague	13	37849	22.0	0.0

Independent Research

Best Paper

Neural Network Analysis of Event-related Potential (ERP) Data

David L. Ryan-Jones and Gregory W. Lewis

Abstract

Cortical event-related potentials (ERPs) reflect sequences of information processing in different areas of the brain. ERPs are generated by both linear and nonlinear processes, and as a result, may be difficult to analyze. In this research, the backpropagation network algorithm was applied to analyses of ERP data. The results suggested that neural network techniques offer an improved and practical alternative to traditional statistical methods.

Introduction

Recent research has suggested that cortical information processing involves the activation of complex serial and parallel neural networks (Van Essen, Anderson, & Felleman, 1992). These networks may span a large number of discrete cortical areas, and these areas may be active for several seconds after a processing sequence is initiated. The actual sequence and location of processing can often be inferred from the physiological activity which results from the processing (Posner, Petersen, Fox, & Raichle, 1988). This is because the temporal sequence of brain processing at any cortical location is presumed to be preserved in the neuroelectric and neuromagnetic variations recorded at the scalp above that location (Nunez, 1981). As a result, it may be possible to relate individual differences in patterns of cortical activity to differences in skill and ability.

There are several potential problems in utilizing an electrophysiological measure, such as the cortical event-related potential (ERP) to predict behavior. First, there are several ERP waveform components and measures that could be used for behavioral prediction. Each component may be effected in a different manner by stimulus, task, anatomical, psychological, and physiological characteristics (Roemer, Josiassen, & Shagass, 1990). Second, there are large differences between subjects in the ERP waveform features elicited by any particular stimulus. Some normal subjects may not show all of the known waveform components, or there may be differences in the

relative timing of these components. Presumably, this reflects
real differences in sensory, cognitive, and motor processes.
Third, cortical processing is rarely restricted to one cortical area,
and some ERP components may be generated independently in
more than one cortical area (Ruchkin, Johnson, Canoune, Ritter,
& Hammer, 1990). Fourth, ERP waveform components are
known to be the result of nonlinear electrochemical processes,
and in general are not the result of simple recruitment of neurons
at a single cortical site. Figure 1 shows a schematic ERP
waveform lasting about 700 msec. Sensory, cognitive, and
motor processing are often contained in a single ERP record, as
suggested in this figure.

EVENT RELATED POTENTIAL (ERP)
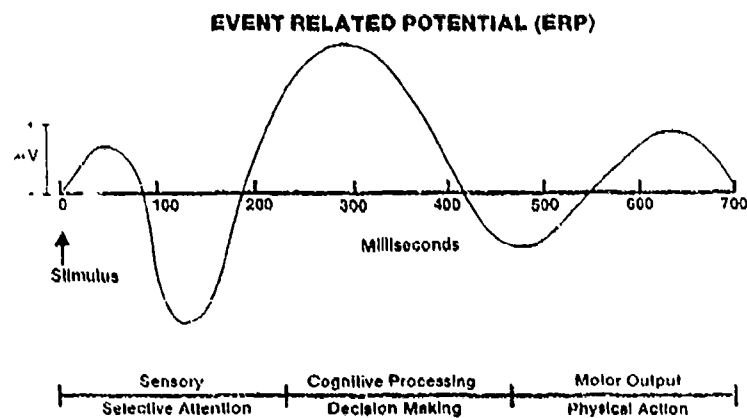
Figure 1. Schematic ERP waveform.

The traditional parametric statistical methods used for prediction
and classification, such as multiple regression and discriminant
analysis, are often used with ERP data. Their use under normal
conditions is limited to linear trends in data that meet the
underlying assumptions of these techniques. Data
transformation can be a useful technique in the analysis of
nonlinear trends by these methods, but it may prove difficult to
capture the essence of the relationships if the true nature of the
underlying function is not known. Most of the commercially
available statistical packages have a limited selection of
nonparametric methods and limited capability for handling
nonlinear datasets. In summary, these methods may not be the
optimal way to develop predictors or classifiers when the data
are nonlinear and nonparametric.

One technology that has been used in recent years to deal with complex datasets is the artificial neural network. An artificial neural network can be described as a hardware or software model that mimics the computational ability of real neurons, such as learning and memory (Maren, 1990). There are many different types of network algorithms, and each can be described in terms of the input and output characteristics, number of layers, nodal connections, training method, and learning parameters (Bailey & Thompson, 1990). Artificial neurons or nodes can be either fully or partially interconnected to other layers of nodes, and these interconnections may be forward, backward, or lateral, depending upon the architecture of the network. Information that is passed to another node may be transformed and weighted. It is typically the values of the connection weights that are modified during the learning process. Network methods can work well with linear input-output relationships, and nonlinear relationships can be captured by a variety of transformation functions, including logistic, gaussian, and step functions. Neural networks can "learn" the correct associations between input and output by one of several self-modification algorithms or learning rules (Zeidenberg, 1990).

Neural network algorithms have been successfully applied to a variety of real-world problems, including feature classification, adaptive control, signal filtering, image compression, neural modeling, evaluation of bank loans, airline flight scheduling, and production control. In general, the network solutions to these applications were shown to work at least as well as solutions derived from statistical methods or expert judgment. One advantage of neural network techniques is that they do not require many assumptions about the nature of the relationship between the predictor variables and the behavioral criteria. One of the most widely used and best understood network algorithms is the backward error propagation, or simply backpropagation network (Wasserman, 1989). The relationship between backpropagation and more traditional statistical techniques has been reviewed (White, 1989a). Learning in a backpropagation network with continuous input data, logistic data transformation, and only input and output layers is analogous to nonlinear multiple regression. The two techniques are best viewed as alternative statistical methods for solving the least squares problem (White, 1989b). It has also been shown that backpropagation networks can estimate the optimal Bayesian discriminant function, and, under certain conditions, the output of the network represents the a posteriori probability of class membership (Wan, 1990; Shoemaker, 1991).

A backpropagation network can be described as a feedforward network with backward adjustment for error in the connection weights. The network must consist of at least two layers, but in a more typical network there are three or more layers: an input layer, one or more hidden layers, and an output layer. Data passed from node to node can only flow in the forward direction (i.e., input to output), and each node may be connected to every other node in the next layer. The weights of the connections are generally randomized at the beginning of training. Training data must specify the output expected from the network as a result of the input data. During training, the weights of the connections between nodes are modified as a function of the degree of error between the actual and expected outputs. Weights are first modified at the output layer and then in each previous hidden layer (i.e., backward propagation). Once trained, network generalization may be evaluated with a testing dataset containing examples that are different than the examples in training dataset (Dayhoff, 1990). A schematic diagram of a feedforward network is shown in Figure 2. This figure shows four nodes in the input layer, three nodes in the hidden layer, and one node in the output layer.

The objective of this project was to assess the utility of the backpropagation network in the analysis of ERP data for three real-world applications. The backpropagation algorithm was

OUTPUT

HIDDEN

INPUT
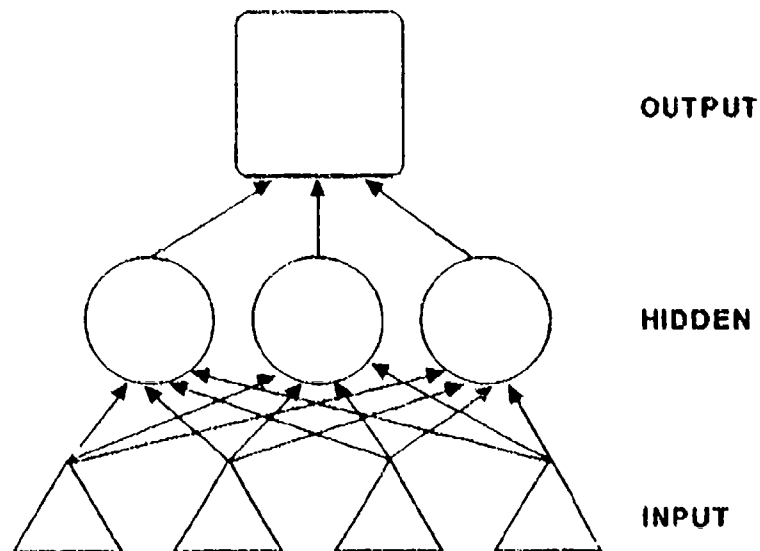
Figure 2. Schematic artificial neural network.

selected for this project because of its success in previous application studies, commercial availability of the software, and the high level of understanding of the workings of the algorithm in the scientific community. Specifically, the areas of interest for the project were: (1) real-time performance monitoring and control in human-machine systems, (2) identification of individuals from brainwave patterns, and (3) prediction of rifle marksmanship from performance on a laboratory simulation task. In the studies that follow, a commercially available backpropagation network software package running on a 20 MHz, 386-based, IBM compatible personal computer was used.

## Real-time Performance Monitoring

Real-time adaptive control of machine activity, based upon behavioral and physiological responses from the human in the system, has been an elusive goal of engineering and psychology. ERP data may be considered as one category of information processing measures, which could potentially be utilized in an adaptive control system. However, there are several problems that must be solved before such a system could be considered to be practical and reliable. ERP data frequently have a low signal-to-noise ratio. This means that several samples of the signal usually must be averaged together to reduce the amplitude of uncorrelated noise. ERP data may also contain a variety of artifacts, such as eyeblinks and muscle activity. Since the amplitude of these artifacts may exceed the amplitude of the ERP, a real-time monitoring system must be able to distinguish these artifacts from the stimulus-evoked neural activity. In an adaptive control system, it is important to be able to determine the state of the operator at any instant, perhaps from a single sample of the signal. In the following examples, the ability of a backpropagation network to extract artifact and signal information from single samples (epochs) of ERP data was evaluated.

Eyeblink Artifact Detection

Artifacts are common in ERP data (Ryan-Jones & Lewis, 1991b). Eye movements, eyeblinks, and other muscle activity are the most common sources of artifacts seen in data recorded from the scalp. Eyeblink artifacts are generally the largest in amplitude. These artifacts are much larger (i.e., millivolts) than the typical ERP signal (i.e., microvolts), and must generally be

removed from the data before processing. The common approach is to exclude any epoch with artifacts from subsequent analyses. Automated rejection algorithms may identify as many as 70 percent of all eyeblink artifacts, but may also incorrectly identify large ERP features as artifacts. This error may be due in part to the large differences between subjects in the characteristics of artifacts and waveform features. The only sure way to identify all eye movement artifacts is to monitor eye movement activity with a dedicated set of electrodes. This would not be very practical in an operational environment. The purpose of the current analysis was to determine if a backpropagation network could be trained to recognize eyeblink artifact patterns in single-epoch data recorded from standard electrode sites. The data used in this, and some of the following analyses, were extracted from an existing ERP database collected from 130 Marines at Camp Pendleton, CA. The Marines performed a 400 trial, two-letter visual discrimination task over a 24 minute period. ERP epochs, sampled at 128 Hz, were recorded on each trial from 200 msec prestimulus to 1,000 msec poststimulus. Recordings were made from the mid-line frontal (anterior association, Fz), central (sensory-motor, Cz), parietal (posterior association, Pz), and from both hemispheres in the occipital (vision, O1 and O2) regions.

This analysis used data from five subjects randomly selected from a subgroup of subjects with eyeblink artifacts on 30-50 percent of the trials. Previous experience had shown that backpropagation networks with more than 100 input elements, and more than a few hundred training examples can take weeks or months to approximate the underlying relationships in the data. Therefore, in order to reduce training time, not all of the available subjects were selected for this analysis. The data consisted of the 1,000 msec poststimulus interval for each of the 400 epochs for each subject recorded from the most anterior electrode site (Fz). This recording site had the largest eyeblink artifacts, compared to the other sites, due to its proximity to the eyes. Out of the 2,000 epochs, 1,000 odd-numbered epochs were used to train the network, and 1,000 even-numbered epochs were used to test the network. This strategy was adopted because there are often significant changes in ERP amplitude, waveform morphology, and eyeblink frequency over a long recording period, as well as large individual differences in waveform morphology. Therefore, both the training and testing data in this study could be expected to have similar characteristics. Sample artifact data are shown in Figure 3. Eyeblink artifacts are seen in epoch numbers 281 and 288, a large baseline artifact (negative shift) is seen in epoch number 292, and muscle artifact is seen in epoch number 293.
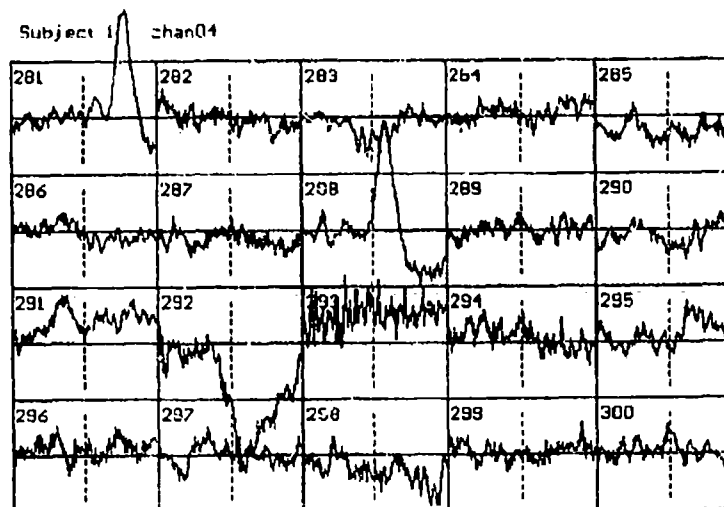
Figure 3. Sample single epoch data for the FZ electrode site.

This study utilized a three-layer network consisting of an input layer of 128 nodes (the 128 post-stimulus data points from Fz), one hidden layer with 128 nodes, and an output layer with 1 node (artifact-no artifact). The input data were normalized to a value between 0 and 1, and the output was allowed to range from 0 to 1. The network was trained with a learning rate of 1.0, and a logistic function was used as the data transform. The network was repeatedly exposed to the training examples (epochs) until all of the examples were correctly classified in a single pass through the data. Correct classification was considered to be an output value of 0.9 or greater for the correct category and 0.1 or less for the incorrect category (100% correct). During testing, these values were 0.51 and 0.49, respectively. Training required about 7 hours of computer time and 242 passes through the training dataset. The trained network was then able to correctly classify 95 percent of the epochs in the test set ($X^2 = 705.07$, $df = 1, p < 0.0001$).

Stimulus Identification

The purpose of this analysis was to determine whether a backpropagation network could learn to identify the category of a visual stimulus presented to a subject from the cortical evoked potential elicited by the stimulus. In other words, it may be possible to tell what stimulus category is being viewed by an individual based upon the resulting cortical activity. The ERP

waveform data for this application came from five subjects who were asked to discriminate between 40 geometric designs and 40 human faces. During performance of the task, ERP epochs, sampled at 128 Hz, were recorded from 125 msec prestimulus to 825 ms poststimulus at sites F3, F4, C3, C4, P3, P4, O1, and O2. In order to reduce the dataset to a more reasonable size, this analysis utilized the data from one pair of sites (P4, F3). These two sites were selected because of previous studies of hemispheric differences in the processing of facial stimuli. The 200 odd-numbered epochs were used to train the network and the 200 even-numbered epochs were used to test the trained network.

The backpropagation network consisted of an input layer with 256 nodes (128 data points each from P4 and F3), one hidden layer with 128 nodes, and an output layer with 1 node (face-not face). Input values were normalized to values between 0 and 1 and the output value was allowed to vary between 0 and 1. The network was trained using a logistic transform function and a learning ate of 0.7. Training continued until all epochs were correctly classifie during a single pass through the data. During training, correct classification was considered to be an output of 0.9 or above if the stimulus was a face, or an output of 0.1 and below if the stimulus was a geometric figure. During testing, these values were 0.51 and 0.49 spectively. The training required about 72 hours of c er time and 2,000 passes through the training set. Th ned network was then able to correctly classify 178 (89 he epochs in the test set ($X^2 = 122.12$, df = 1. $).06. \.

Response Predi:

The objective of this analysis was to determine if a back-propagation network could be trained to predict whether or not a subject would make a correct or an incorrect behavioral response. It is possible to distinguish between correct and incorrect responses from average ERP data using traditional statistical techniques. However, the relatively low signal-to-noise ratio in the single epoch data effectively prevents letermination of the accuracy of the response at this level. Since these data were to be typical of what could be expected in a real-time environment, the study did include ERP data containing eyeblink artifacts. The datab e used in this analysis consisted of the poststimulus portior he epochs from the same five subjects used in eyeblink tion analysis. In addition, a flag value was added to each epoch to indicate

whether or not the epoch contained an artifact. Again, the training set consisted of the 1,000 odd-numbered epochs and the test set consisted of the 1,000 even-numbered epochs.

The backpropagation network consisted of an input layer with 129 nodes (128 poststimulus data points, and 1 artifact flag), one hidden layer with 128 nodes, and an output layer with 3 nodes (hit, miss, or no response). Input data were normalized to values between 0 and 1 and the output values were allowed to range from 0 to 1. The network was trained with a learning rate of 0.7 and used a logistic data transform. Training continued until all the epochs in the training set were correctly classified during a single pass through the data. Correct classification was considered to be an output of 0.9 and above for the correct category and 0.1 and below for the two incorrect categories. During testing, these values were 0.51 and 0.49, respectively. The training required about 48 hours and 360 passes though the data. The trained network was then able to correctly classify 925 (93%) of the epochs ($X^2 = 1036.04$, df $= 4$, $p < 0.0001$).

Individual Identification

Individual identification is an essential part of any security access control system. Neuroelectric and neuromagnetic measures may offer a new method for access control. Although the shape of an ERP waveform differs greatly from subject to subject, there is considerable stability in ERP waveform shape over time within subjects (Lewis, 1984). Therefore, it may be possible to use ERP data for individual identification (Lewis & Ryan-Jones, 1992). The purpose of this study was to determine whether or not backpropagation network could be trained to identify individuals from their ERP waveform pattern.

The ERP data, which were used in this analysis, came from 35 subjects who performed the two-letter discrimination task described in the previous section. The 400 epochs of ERP data for each subject from site Pz were divided into eight successive blocks of 50 epochs. Within each block of 50 epochs, the first 10 epochs that were artifact-free and correct in response were averaged within the block to produce a total of eight ERPs for each subject. Each ERP was then divided into 25 windows, with each window containing 6 sampling points. The points in each window were then averaged to produce a single value for each window. These windowed ERPs were then used as the database for the neural network analysis. The purpose of windowing these

data was to reduce the time required to train the network by reducing the size of the network.

In all, there were eight ERPs for each of the 35 subjects (i.e., one ERP for each trial block). The four odd-numbered block ERPs for each subject were used for training, and the four even-numbered block ERPs for testing. Each dataset contained a total of 140 ERPs (35 subjects x 4 block ERPs). The backpropagation network consisted of an input layer of 25 nodes (ERP windows), one hidden layer with 25 nodes and an output layer with 35 nodes (subjects). Input data were normalized between 0 and 1 and the values of the output nodes were allowed to vary between 0 and 1. The network was trained with a logistic data transform and a learning rate of 1.0. Training continued until the network was able to correctly classify all of the ERPs in the training set in a single pass. During training, correct classification was considered to be an output value of 0.9 or greater for the correct category and 0.1 or less for all of the other categories. During testing, the values were 0.51 and 0.49, respectively. Training required about 5 hours and 700 passes through the data. The trained network was then able to correctly classify 73 (52%) of the ERPs in the testing set ($p < 0.0001$). Statistical significance was assessed by simple probability (1 chance in 35 of correct assignment on any trial).

One useful technique for the evaluation of network generalization is the comparison of the relative amplitudes of the outputs. During testing, more than one output node may generate an output value. In this analysis, there were relatively large output values for up to about six output categories. In some cases, the correct output category may not have the largest amplitude, but instead may have the second or third highest value. Although the network may not correctly identify the individual, it can still narrow identification to two or three individuals. For some purposes, this could still be acceptable performance. When the second and third highest output values (out of 35 possible outputs) were examined, it was found that the correct category had the second highest value on 25 trials and the third highest value on 15 trials. In summary, on 113 of the 140 trials (81%), the ERPs produced the highest, second highest, or third highest output value.

Prediction of Marksmanship Performance

Rifle marksmanship has long been a subject of intense study in military psychology. Two important components of marksmanship are visual search and visual discrimination. Previous research suggested that measures of these visual factors correlate with rifle qualification scores and that ERP measures correlate with visual performance. The purpose of this evaluation was to determine if a backpropagation network could predict qualification category from ERP data collected during the two-letter visual discrimination task described in a previous section (Ryan-Jones & Lewis, 1991a).

The data used in this analysis were from 67 of the subjects and consisted of the epochs recorded from sites O1 and O2. These sites are located over the visual areas of the cerebral cortex and the ERPs were expected to be associated with sensory processing of the letters presented during the task. The database was created by averaging the first 20 artifact-free epochs with a correct response in each of 8 successive blocks of 50 trials, for a total of 160 epochs in each ERP average. Each ERP was then divided into 25 windows by the method described in a previous section. The ERPs from 34 of the subjects were used for training and ERPs from the other 33 subjects were used for testing.

The backpropagation network with one input layer with 50 nodes (25 windows for each of two ERPs), one hidden layer with 50 nodes, and an output layer with 1 node (expert-not expert). Input data were normalized with a range of 0 to 1 and the output value was allowed to range form 0 to 1. The network was trained with a logistic data transform and a learning rate of 0.8. Training continued until all of the ERPs in the training set were correctly classified during a single pass. Classification was considered to be correct if the output value was 0.9 and above for the correct category and 0.1 or less for the incorrect category. The values for testing were 0.51 and 0.49, respectively. Training required about 1 hour and 60 passes through the dataset. The trained network was then able to correctly classify 23 of the 33 subjects (72%) in the test set ($X^2 = 4.95$, df = 1, $p < 0.05$). Network performance was compared with a discriminant analysis of the same dataset and performance of the discriminant function was at chance level ($X^2 = 1.51$, df = 1, $p > 0.05$).

Conclusions

During this project, the utility of a backpropagation neural network in the analysis of ERP data was assessed in three application areas. In each area, neural network techniques do appear to be useful in the analysis of ERP data. There are many positive aspects about these techniques. First, the user does not have to make assumptions about the data other than to identify all of the possible variables that could predict the criteria. Second, a neural network approach may be better than traditional statistical methods for classification if there are complex nonlinearities in the data. A neural network with 1-3 hidden layers can represent almost any function in a dataset, capturing both the linear and nonlinear aspects of the data. Finally, neural networks require minimal preprocessing of ERP data. This means that when appropriate, raw ERP data can be the input into the network. Once a network is trained, new data can be classified in milliseconds. This greatly improves the likelihood that ERP data can be collected and analyzed in near real-time on adaptive control tasks.

There are several potential drawbacks in using neural network methods. First, selection of the appropriate network algorithm and design of the network may be a problem. Selection of the algorithm is constrained by the database characteristics and the projected use of the network. Design of the network can require a great amount of experimentation to determine the optimal structure. Second, the learning process may take a relatively long period of time. If there are more than 100 input variables and thousands of examples, training may literally take days or weeks. Third, very minor changes in the design of the net and the initial conditions can make a great deal of difference in the rate of convergence and in the final level of training and generalization. Unacceptable training or testing performance may only mean that the network structure is suboptimal, or that the initial conditions need to be changed. Finally, interpretation of the structure and weighting of a trained network may be very difficult. Weights can depend upon both the importance of a connection and the range of the values passed over the connection. This may place some limits on the usefulness of neural networks for exploratory data analysis. Recently, new research has suggested several useful strategies for the interpretation of connection weights (Garson, 1991). Even with the potential problems suggested here, there are many potential applications of neural network techniques that this Center will continue to explore in the future.

References

Bailey, D., & Thompson, D. (1990). How to develop neural-network applications. *AI Expert, 6*, 38-48.

Dayhoff, J. (1990). *Neural network architectures: An introduction*. New York: Van Nostrand Reinhold.

Garson, G. (1991). Interpreting neural-network connection weights. *AI Expert, 4*, 47-51.

Lewis, G. (1984). Temporal stability of multichannel, multimodal ERP recordings. *International Journal of Neuroscience, 25*, 131-144.

Lewis, G., & Ryan-Jones, D. (1992, June). *Neural network identification of individuals from ERP patterns*. Paper to be presented at the Fourth Annual Meeting of the American Psychological Society, San Diego, CA.

Maren, A. (1990). Introduction to neural networks. In A. Maren, C. Harston, & R. Pap (Eds.), *Handbook of neural computing applications*. San Diego: Academic Press.

Nunez, P. (1981). *Electric fields of the brain: The neurophysics of EEG*. New York: Oxford University Press.

Posner, M., Petersen, S., Fox, P., & Raichle, M. (1988). Localization of cognitive operations in the human brain. *Science, 240*, 1627-1631.

Roemer, R., Josiassen, R., & Shagass, C. (1990). Comparing principal components analyses of evoked potentials recorded from heterogeneous groups of subjects. *Psychophysiology. 27*, 101-110.

Ruchkin, D., Johnson, R., Canoune, H., Ritter, W., & Hammer, M. (1990) Multiple sources of P3b associated with different types of information. *Psychophysiology, 27*, 157-176.

Ryan-Jones, D., & Lewis, G. (1991a). Application of neural network methods to prediction of job performance. *Proceedings of the Second Workshop on Neural Networks: Academic/Industrial/NASA/Defense* (pp. 731-736). Madison, WI: Omnipress.

Ryan-Jones, D., & Lewis, G. (1991b). Application of neural network techniques to artifact detection and correction in cortical event-related potentials. *Proceedings of the Second Annual R&D Information Exchange Conference* (NWC TS 91-22, p. 57). Naval Weapons Center, China Lake, CA.

Shoemaker, P. (1991). A note on least-squares learning procedures and classification by neural network models. *IEEE Transactions on Neural Networks, 2,* 158-160.

Van Essen, D., Anderson, C., & Felleman, D. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science, 255,* 419-422.

Wan, E. (1990). Neural network classification: A Bayesian interpretation. *IEEE Transaction on Neural Networks, 1,* 303-305.

Wasserman, P. (1989). *Neural computing: Theory and practice.* New York: Van Nostrand Reinhold.

White, H. (1989a). Learning in artificial neural networks: A statistical perspective. *Neural Computing, 1,* 425-464.

White, H. (1989b). Neural-network learning and statistics. *AI Expert, 12,* 48-52.

Zeidenberg, M. (1990). *Neural Networks in artificial intelligence.* New York: Ellis Horwood.

Best Paper Nomination Rationale

Research Merit

Event related brain potentials (ERPs) are seeing greatly increased use as research tools, in addition to clinical applications, for assessing sensory/perceptual, other physiological functions, and cognitive (decision making) processes. To date, complexity of ERP signal processing and analysis requirements has limited the use of ERPs for many applications, including real-time monitoring in military environments. Neural network (NN) procedures offer potential new power to ERP signal processing and analytic areas. Even though ERP and NN research and development has been done for several years, the use of NN technology for assessing ERP records of cognitive processing is a very recent application.

Research Approach/Plan/Focus/Coordination

The approach of this research was to demonstrate the application of backpropagation neural networks in three potential application areas of ERP data. These included:

1. Real-time monitoring in operational systems, including target recognition and behavioral response prediction.

2. Individual identification in personnel security.

3. Rifle marksmanship performance prediction.

Coordination has been maintained with the neural network research community, as well as the ERP and applied research communities. The relevant research literature has been monitored constantly.

Difficulty of Problem Addressed

Effective prediction of human performance has remained a desirable, but intractable, goal in personnel assessment. Current assessment techniques account for a reliable, but relatively small, portion of the variance in predicting job performance. This has remained so although large amounts of manpower and funding have been expended over many years on research intended to improve prediction. The new capability of assessing

cognitive processing, including decision-making and workload through direct brain recordings (ERPs), shows greater promise than other assessment techniques for improving performance prediction. Cognitive processing is extremely complex even for relatively simple tasks. Complexity of ERP preprocessing, signal processing, analyses, and subsequent application to operational problems is great. But, it no longer remains intractable, due to powerful new tools such as neural network analyses, ERP recordings, and brain imaging and mapping techniques.

Originality of Approach

We know of no other application of neural network techniques to ERP analyses for personnel assessment and performance prediction.

Potential Impact on Navy/Center Needs

There is increased recognition of the need for improved assessment of cognitive processing during the performance of most military tasks, especially those found in Combat Information Center environments, (e.g., EW, ASW). With improved assessment of personnel decision-making requirements and capabilities, large reductions in personnel and training costs will be realized. The current research has demonstrated several points of impact on Navy needs, including (1) research has shown that it may be possible to extract and use information from ERP data as feedback in adaptive training and in operational systems, and (2) research has shown that neural network techniques can identify meaningful relationships between ERP and performance data, which may not be obtained by traditional statistical techniques.

Probability of Achieving Impact on Naval Needs

There is moderate-to-high risk and high payoff for improved personnel assessment, training, and job performance prediction using NN and ERP technology. Prognosis is good and probability is increasing that these two technologies will effect Navy personnel needs in a positive way.

Publication and Presentations

The following papers report the results of this effort.

Ryan-Jones, D., & Lewis, G. (1991a). Application of neural network methods to prediction of job performance. *Proceedings of the Second Workshop on Neural Networks: Academir· dustrial/NASA/Defense* (pp. 731-736). Madiso· vI: Omnipress.

Ryan-Jones, D., & Lewis, G. (1991b). Application of neural network techniques to artifact detection and correction in cortical event-related potentials. *Proceedings of the Second Annual R&D Information Exchange Conference* (NWC TS 91-22) (p. 57). Naval Weapons Center, China Lake, CA.

Lewis, G., & Ryan-Jones, D. (1992, June). *Neural network identification of individuals from ERP patterns.* Paper to be presented at the Fourth Annual Convention of the American Psychological Society, San Diego, CA.

Ryan-Jones, D., & Lewis, G. (in press). Neural network and ERP technologies: Implications for military task and job performance assessment. *Journal of Military Psychology.*

Independent Exploratory Development

Best Project

Effects of Administration Method and Anonymity/Identification on Survey Responses

Stephanie Booth-Kewley, Jack E. Edwards, and Paul Rosenfeld

Abstract

Researchers have sought to reduce socially desirable responding on self-report instruments. A number of studies have claimed that computer-administered attitude and personality assessments result in less socially desirable responding than do paper-and-pencil assessments. Lautenschlager and Flaherty's (1990) recent study unexpectedly found that more socially desirable responding occurred on computers. The present study attempted to replicate and extend their findings. Male Navy recruits (N = 246) completed several questionnaires including the Balanced Inventory of Desirable Responding (Paulhus, 1991), a measure of impression management and self-deceptive enhancement. Respondents completed the questionnaires in either a computer-administered or paper-and-pencil condition, and were either Anonymous or Identified. The results supported Lautenschlager and Flaherty's (1990) finding that Identified respondents had higher impression management and self-deceptive enhancement scores than Anonymous respondents. Contrary to their results, there was no systematic difference between computer and paper-and-pencil modes. It is concluded that computer and paper-and-pencil modes of administration yield similar responses on attitude questionnaires.

Introduction

Computer administration of psychological and organizational instruments is becoming increasingly widespread (Bertram & Bayliss, 1984). With this increased usage has come the idea that computer administration might reduce or counteract social desirability bias--the tendency "to stretch the truth in an effort to make a good impression" (Martin & Nagao, 1989, p. 72). On self-report attitude and personality instruments (e.g., psychological tests, surveys, and questionnaires), social desirability is seen by some observers as pervasive and problematic (Murphy & Davidshofer, 1991). Demonstrating that computerized self-report attitude and personality instruments elicit less social desirability bias than their paper-and-pencil counterparts would argue for greater reliance on computer

administration, given computerized administration's many other advantages (e.g., ease of administration, elimination of missing responses) elimination of data entry, automatic scoring, and item-branching capabilities (Rosenfeld, Doherty, & Carroll, 1987).

In studies comparing paper-and-pencil and computer-administered attitude and personality questionnaires, the findings have been equivocal. Studies comparing these two administration modes have found that respondents completing questionnaires on a computer admitted more anxiety symptoms and scored lower on a lie scale (Evan & Miller, 1969), reported more fear (Carr & Ghosh, 1983), gave fewer socially desirable responses (Kiesler & Sproull, 1986; Martin & Nagao, 1989), and were less likely to over-report their SAT scores (Martin & Nagao, 1989). However, other studies (Erdman, Klein, & Griest, 1983; Kantor, 1991; Katz & Dalby, 1981; Lukin, Dowd, Plake, & Kraft, 1985; Millstein, 1987; Rozensky, Honor, Rasinski, Tovian & Herz, 1986; Skinner & Allen, 1983; White, Clements, & Fowler, 1985; Wilson, Genco, & Yager, 1985) have found that these two administration modes yielded very similar results. For example, Rosenfeld, Doherty, Vicino, Kantor, and Greaves (1989) found that although respondents enjoyed completing a computerized attitude questionnaire more than a paper-and-pencil version, the means for the decision-making scale were not significantly different. Three recent studies (Davis & Cowles, 1989; Lautenschlager & Flaherty, 1990; Schuldberg, 1988) further muddle the situation by reporting that individuals who answered questionnaires using a computer emitted more socially desirable responses and were less candid than peers using paper-and-pencil. More specifically, Davis and Cowles (1989) reported that individuals in the computer condition admitted fewer anxiety symptoms than those in the paper-and-pencil condition. Schuldberg (1988), similarly, found that individuals in the computer condition reported relatively less psychopathology than did people in the paper-and-pencil condition.

In the most recent study comparing paper-and-pencil versus computerized attitude questionnaires, Lautenschlager and Flaherty (1990) had college students complete Paulhus's (1984) Balanced Inventory of Desirable Responding (BIDR). The BIDR contains two scales: impression management (IM) and self-deception (SD), that assess socially desirable responding. IM is the deliberate tendency to over-report desirable behaviors and under-report undesirable ones; whereas, SD is the tendency to give honest but overly positive reports about one's self. SD

differs from IM in that the respondents actually believe their positive SD self-reports (Paulhus, 1984). Contrary to their predictions, Lautenschlager and Flaherty (1990) found that students using computers gave more socially desirable responses on both BIDR scales than did students completing a paper-and-pencil version. Also, students whose responses were identified had significantly higher scores on both scales than did students in the anonymous condition. These results led the authors to conclude that computer administration of psychological instruments may increase the tendency to engage in impression management.

While researchers (e.g., Kiesler & Sproull, 1986; Rosenfeld et al., 1991) have speculated as to why computers might reduce socially desirable responding on attitude and personality instruments (e.g., the computer creates an impersonal social setting), it is unclear why computer administration of such questionnaires would reduce candor and increase socially desirable responding. An explanation offered by Lautenschlager and Flaherty (1990) pertained to the way in which the attitude questionnaires in their computer condition were administered. In their computer condition, respondents had to answer each item before the next item would appear, and they could not look at or change earlier answers. The authors suggested that having the respondents complete the attitude questionnaires in this "lock-step" fashion may have caused them to engage in more impression management than would have occurred if respondents had been able to "backtrack."

Lautenschlager and Flaherty's (1990) findings pose a challenge to the continued administration of computerized attitude and personality questionnaires. As they noted, "the administration of ... attitude questionnaires in organizational research may be adversely affected when converted from paper-and-pencil format. Increases due to impression management on such diagnostic measures may produce inaccurate and potentially misleading results" (p. 314). Given that organizations will probably increase their reliance on computers for administering self-report instruments, there is a clear need to determine whether responses on computerized attitude and personality instruments are systematically different from those obtained with paper-and-pencil.

With this overall goal in mind, the present study compared responses of Navy recruits on the IM and SD scales using Paper-and-pencil, Computer Backtrack, and Computer No-backtrack

conditions. Approximately half of the recruits answered the questionnaire anonymously, while the other half identified themselves.

Method

Subjects

Navy recruits (N = 246) completing basic training in San Diego, CA served as survey respondents. Because only male recruits receive basic training in San Diego, females were not included in this study. The respondents were predominantly nonHispanic whites (78%), with small proportions of Hispanics (10%), blacks (6%), and individuals from other ethnic groups (6%). Each respondent had a high school diploma or general equivalency degree; 41 percent had also completed some college. The majority (89%) of the respondents were single. The mean age was 20 years.

Measures

All respondents provided their age, race/ethnicity, education, and marital status. Respondents answered all nondemographic items with ratings that varied from *Strongly Disagree* (1) to *Strongly Agree* (5).

The IM and SD Enhancement (SDE) scales of the BIDR Version 6 (Paulhus, 1991) were administered. Sample items from the 20-item IM scale include "I always obey laws, even if I'm unlikely to get caught" and "I never cover up my mistakes." The SDE scale is a newer version of the BIDR-SD scale. Sample items from the 20-item SDE scale include "I never regret my decisions" and "My first impressions of people usually turn out to be right." Higher scores indicate greater impression management or self-deceptive enhancement. Paulhus (1991) found internal consistency reliabilities ranging from .75 to .86 for IM and from .68 to .80 for SDE.

The 15-item Organizational Commitment Questionnaire (Mowday, Steers, & Porter, 1979) was adapted for Navy use by substituting "the Navy" for "this organization." Mowday et al. (1979) reported internal consistency reliabilities of .82 to .93 for their scale. Although Lautenschlager and Flaherty (1990) did not administer this scale, it was included in the present study

Effects of Administration Method and Anonymity/Identification on Survey Responses

because of its relevance to industrial/organizational psychologists. It was of interest to determine whether organizational scales that applied researchers might use would be affected by administration mode.

Respondents also completed the 19-item Computer Anxiety Rating Scales (CARS), a measure of computer anxiety and negative attitudes toward computers (Heinssen, Glass, & Knight, 1987). Heinssen et al. reported a coefficient alpha of .87 for their measure. The CARS was administered to make sure that respondents in the computer and paper-and-pencil conditions did not differ in computer anxiety. Such a difference would make it impossible to determine whether differences in IM/SDE scores are due to the computer versus paper-and-pencil manipulations, pre-experimental differences on computer anxiety among the respondents, or both.

Respondents provided information that was used to check the manipulations in this study. In a postexperimental questionnaire, respondents answered items (specified in the Results) that assessed how anonymous they believed that their responses were, how important and interesting they thought the questionnaires were, and how they felt while completing them.

Design and Procedure

Anonymity level and administration mode were the two independent variables. There were two levels of anonymity: Anonymous (i.e., explicitly told to avoid indicating name or social security number) and Identified (i.e., explicitly told to indicate name and social security number). The three levels of administration mode were Computer Backtrack, Computer No-backtrack, and Paper-and-pencil. In the Computer Backtrack conditions, respondents were allowed to backtrack within the questionnaire and change previous answers. Respondents in the Computer No-backtrack condition could not backtrack within the questionnaire. Respondents using paper-and-pencil could, of course, change previous responses.

Respondents in all experimental conditions were administered the questionnaires in groups of 10 to 20 people, as a scheduled part of basic training. The respondents were randomly assigned to one of six conditions: Anonymous Paper-and-pencil ($N = 42$), Identified Paper-and-pencil ($N = 42$), Anonymous Computer Backtrack ($N = 40$), Identified Computer Backtrack ($N = 40$), Anonymous Computer No-backtrack ($N = 40$), and Identified

Computer No-backtrack ($N = 42$). Not included in these cell sizes are three recruits, all assigned to Paper-and-pencil conditions who chose not to volunteer for this study.

Respondents completed the demographic items, the BIDR, and the Organizational Commitment Questionnaire in the administration mode (i.e., computer or paper-and-pencil) to which they had been assigned. After these data were gathered, the CARS and the postexperiment questionnaire were administered to everyone in paper-and-pencil format.

The sample and cell sizes for the present study were virtually identical to those of Lautenschlager and Flaherty (1990). A power analysis indicated that the power of the present study for detecting an effect the size of that found by Lautenschlager and Flaherty ($d = .36$) for IM with an .05 alpha was .93 (Cohen, 1977). (A power analysis was not calculated for SDE because Lautenschlager and Flaherty had not predicted an administration mode effect for this variable.)

BIDR Scoring

Paulhus (1990) has recommended two scoring methods for the BIDR: (1) a continuous scoring method in which items are simply summed after negatively-keyed items have been reverse scored and (2) a dichotomous scoring method in which one point is awarded for each "extreme" response (i.e., "6" or "7" for positively-keyed items and "1" or "2" for negatively-keyed items). Because Lautenschlager and Flaherty (1990) used the continuous scoring method and replication of their study was our major goal, we used the continuous scoring method. However, in response to a reviewer's suggestion, we also rescored the BIDR scales dichotomously. Because of the loss of information incurred by dichotomization, we focus on the results from the continuous method but report results for both methods.

Results

Analyses of variances (ANOVAs) were performed to test the effects of administration mode and anonymity level on the SDE and IM scales. These results are presented in Table 1.

Contrary to Lautenschlager and Flaherty's (1990) findings, the main effects of administration mode were *not* significant for either SDE, $F(2, 239) = .33, p > .05$, or IM, $F(2, 239) = .12$,

Table 1

SDE and IM Means and Standard Deviations By Experimental Condition

Scale and Administration Mode	Continuous Scoring						Dichotomous Scoring					
	Anonymity Level						Anonymity Level					
	Anonymous		Identified		Overall		Anonymous		Identified		Overall	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SDE												
Paper-and-pencil	63.33	7.34	63.95	7.65	63.64	7.46	9.80	2.71	9.28	3.09	9.54	2.89
Computer Backtrack	60.87	8.52	65.43	8.60	63.15	8.81	8.35	3.36	9.82	3.77	9.09	3.63
Computer No-backtrack	61.10	9.70	64.95	7.60	63.07	8.85	8.45	3.39	9.71	3.16	9.10	3.32
Overall	61.80	8.56	64.77	7.91	63.29	8.36	8.87	3.19	9.61	3.33	9.24	3.28
IM												
Paper-and-pencil	53.27	11.42	54.67	11.11	53.97	11.22	5.72	4.03	5.90	3.92	5.81	3.94
Computer Backtrack	52.80	11.29	56.83	10.69	54.81	11.11	5.27	3.47	6.35	3.76	5.81	3.64
Computer No-Backtrack	53.47	11.17	56.24	11.66	54.89	11.44	5.45	3.81	6.57	4.01	6.02	3.93
Overall	53.18	11.20	55.92	11.11	54.56	11.22	5.48	3.75	6.28	3.88	5.88	3.83

$p > .05$ for the continuous scoring method. The ANOVAs revealed only main effects of anonymity level for both SDE, $F(1, 239) = 7.25, p < .05$ and IM, $F(1, 239) = 3.76, p = .05$. Consistent with Lautenschlager and Flaherty's (1990) findings, Identified individuals presented themselves more favorably than Anonymous respondents on the SDE and IM scales. Also, consistent with Lautenschlager and Flaherty's findings were the nonsignificant administration mode by anonymity interactions for both SDE, $F(2, 239) = 1.65, p > .05$, and IM, $F(2, 239) = .23$, $p > .05$.

When the BIDR scales were scored dichotomously, the overall trends were the same. Once again, the main effects of administration mode were nonsignificant for both SDE, $F(2, 239) = .53, p > .05$, and IM, $F(2, 239) = .07, p > .05$. Similarly, the anonymity level by administration mode interactions were nonsignificant for both SDE, $F(2, 239) = 2.23$,

$p > .05$, and IM, $F(2, 239).39$, $p > .05$. The only difference between results from the continuous and dichotomous scoring methods occurred for the main effects of anonymity level. Although the pattern of scores remained the same, the main effects of anonymity level were not significant for either SDE, $F(1, 239) = 3.18, p > .05$ or IM, $F(1, 239) = 2.58, p > .05$, when dichotomous scoring was used.

In further support of the similarity of results for computer versus paper-and-pencil administration of attitude scales, internal consistency reliability coefficients for the computer versions (collapsed across Backtrack and No-backtrack conditions) of the continuously scored IM and SDE scales (alphas of .85, and .71, respectively) were similar to those found for the paper-and-pencil versions (alphas of .86 and .63, respectively). Finally, the intercorrelations between the two (continuously scored) BIDR scales were .43 collapsed across the two computer conditions and .55 for the paper-and-pencil condition. (The intercorrelations for the Anonymous and Identified conditions were .46 and .44, respectively.) The similar but somewhat lower intercorrelation across the computer conditions contradicted a suggestion made by one of Lautenschlager and Flaherty's reviewers that the increased motivation to appear socially desirable might lead to increased intercorrelations for the computer condition relative to the paper-and-pencil condition.

An ANOVA was performed to test the effects of administration mode and anonymity level on Organizational Commitment. Neither anonymity level, $F(1, 238) = 1.81, p > .05$, nor administration mode, $F(2, 238) = .65, p > .05$, affected commitment scores. The interaction was also not significant, $F(2, 238) = .67, p > .05$.

Computer anxiety (CARS) scores did not vary by anonymity level, $F(1, 241) = .89, p > .05$, or administration modes, $F(2, 240) = 1.11, p > .05$; nor was the interaction significant, $F(2, 239) = .69, p > .05$.

Analyses comparing the postexperiment questionnaire responses for the Anonymous and Identified conditions revealed that Anonymous respondents were less likely to agree that "It would be easy for someone to find out my answers to this survey," $t(243) = 2.10, p < .05$ (one-tailed), and were more likely to agree with the statement, "I doubt that my survey answers will ever be linked with my name or any other information that identifies me," $t(243) = 3.77, p < .05$ (one-tailed).

Analyses comparing the postexperiment questionnaire responses for the computer (collapsed across Backtrack and No-backtrack conditions) and paper-and-pencil conditions showed that respondents in the computer conditions were less likely to think that it would be easy for someone to determine their questionnaire answers, $F(1, 244) = 4.46$, $p < .05$); less likely to think the questionnaire was boring, $F(1, 244) = 8.74$, $p < .05$; and more likely to think it was important, $F(1, 244) = 4.14$, $p < .05$. Also, respondents in the computer conditions were more likely than those in the paper-and-pencil conditions to report feeling very aware of their thoughts and feelings while completing the questionnaire, $F(1, 244) = 4.93$, $p < .05$. Finally, in the Computer Backtrack condition, 61 percent of the respondents backtracked at least once; 45 percent backtracked more than once. The number of times respondents backtracked ranged from 0 to 24.

Discussion

Behavioral and organizational researchers have attempted to reduce social desirability response bias in attitude and personality measurement using methods ranging from the assessment of physiological responses such as pupil dilation (Hess, 1965), galvanic skin response, and facial muscle contractions (cf. Petty & Cacioppo, 1981), to psychological manipulations such as the bogus pipeline (a machine purported to have lie-detecting capabilities [Jones & Sigall, 1971]). In particular, improvements on the basic paper-and-pencil, self-report scale have been eagerly sought. The computer, it appears, has been the latest of these "holy grails." With regard to reducing social desirability bias in attitude and personality measurement, the computer's promise may have been greater than its delivery. The results of the present study suggest that computer and paper-and-pencil modes of administration yield similar results. This finding indicates that, where financial and logistical considerations allow, researchers are justified in using computers instead of paper-and-pencil to gather attitude data because of the many previously cited advantages of computerized data collection.

Our results did not support Lautenschlager and Flaherty's (1990) surprising finding of increased social desirability on the computer. While the main effect of anonymity was replicated for both the IM and SDE scales, there was no significant effect due to mode of administration for IM, SDE, or organizational commitment. Our results also failed to support an explanation

offered by Lautenschlager and Flaherty (1990) for their findings. Lautenschlager and Flaherty suggested that their respondents' inability to backtrack on the computer may have been a reason for their greater impression management (relative to the paper-and-pencil respondents). If true, differences in IM scores would have been expected between the Computer Backtrack and the Computer No-backtrack conditions in the present study. In fact, no difference between these conditions was found.

In trying to explain why their results directly contradicted those of Martin and Nagao (1989), Lautenschlager and Flaherty (1990) pointed to a number of procedural and methodological differences between the two studies. A similar case could be made to explain the results reported in the present study versus the findings of Lautenschlager and Flaherty. For example, they used BIDR Version 3 whereas we used Version 6. Also, their respondents were college students, 70 percent of whom were female, while the respondents in the present study were male Navy recruits. Differences between males and females or between military and college populations could account for our failure to replicate Lautenschlager and Flaherty's results. At any rate, researchers may need to employ less obtrusive measures such as response latency (George & Skinner, 1990) to better determine when and for whom (e.g., see Rosenfeld et al., 1991) computers increase, decrease, or have no effect on socially desirable responding in attitude and personality scales.

Although researchers are often disappointed by null results, there are advantages associated with the present findings of "no difference." Knowing that under certain circumstances computer and paper-and-pencil administered questionnaires yield similar means and standard deviations is valuable for researchers and organizational practitioners faced with the option of giving their questionnaires under one or both of these administration modes. Also, in the present study respondents who used the computer found the questionnaire to be more interesting, more important, and felt more aware of their thoughts and feelings than respondents who used paper-and-pencil. Other researchers (e.g., Lukin et al., 1985; Rosenfeld et al., 1989) have obtained similar results. Thus, when computer-administered attitude questionnaires yield similar scale means, standard deviations, and reliabilities, an argument for the use of computers is that they are typically regarded more favorably by respondents.

In conclusion, the present results of no systematic difference between computer and paper-and-pencil responding on attitude questionnaires are consistent with much of the published research literature. As the use of computers and computer-generated information continues to grow, and as the general population becomes further "enlightened" (Gergen, 1973) about computers' true capabilities, many of the differences that exist between computer and paper-and-pencil administration modes may dissipate even further. For the interim, researchers should attempt to identify the boundary and contextual conditions that produce differences in computer versus paper-and-pencil responses.

References

Bertram, D., & Bayliss, R. (1984). Automated testing: Past, present, and future. *Journal of Occupational Psychology, 57*, 221-237.

Carr, A. C., & Ghosh, A. (1983). Accuracy of behavioral assessment by computer. *British Journal of Psychiatry, 142*, 66-70.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Davis, C., & Cowles, M. (1989). Automated psychological testing: Method of administration, need for approval, and measures of anxiety. *Educational and Psychological Testing, 49*, 311-320.

Erdman, H., Klein, M. H., & Griest, J. H. (1983). The reliability of a computer interview for drug use/abuse information. *Behavior Research Methods and Instrumentation, 15*, 66-68.

Evan, W. M., & Miller, J. R., (1969). Differential effects on response bias of computer vs. conventional administration of a social science questionnaire: An exploratory methodological experiment. *Behavioral Science, 14*, 216-227.

Gergen, K.J. (1973). Social psychology as history. *Journal of Personality and Social Psychology, 26*, 309-320.

George, M. S., & Skinner, H. A. (1990). Using response latency to detect inaccurate responses in a computerized lifestyle assessment. *Computers in Human Behavior*, 6, 167-175.

Heinssen, R. K., Glass, C. R., & Knight, L. A. (1987). Assessing computer anxiety: Development and validation of the computer anxiety rating scale. *Computers in Human Behavior*, 3, 49-59.

Hess, E. H. (1965). Attitude and pupil size. *Scientific American*, 212, 46-54.

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitudes. *Psychological Bulletin*, 76, 349-364.

Kantor, J. (1991). The effects of computer administration and identification on the Job Descriptive Index (JDI). *Journal of Business and Psychology*, 5, 309-323.

Katz, L., & Dalby, J. T. (1981). Computer and manual administration of the Eysenck Personality Inventory. *Journal of Clinical Psychology*, 37, 586-588.

Kiesler, S., & Sproull, L. (1986). Response effects in the electronic survey. *Public Opinion Quarterly*, 50, 402-413.

Lautenschlager, G. J., & Flaherty, V. L. (1990). Computer administration of questions: More desirable or more social desirability? *Journal of Applied Psychology*, 75, 310-314.

Lukin, M. E., Dowd, E. T., Plake, B. S., & Kraft, R. G. (1985). Comparing computerized versus traditional psychological assessment. *Computers in Human Behavior*, 1, 49-58.

Martin, C. L., & Nagao, D. H. (1989). Some effects of computerized interviewing on job applicant responses. *Journal of Applied Psychology*, 74, 72-80.

Millstein, S. G. (1987). Acceptability and reliability of sensitive information collected via computer interview. *Educational and Psychological Measurement*, 47, 523-533.

Mowday, R. T., Steers, R. M., & Porter, L. W. (1979). The measurement of organizational commitment. *Journal of Vocational Behavior*, 14, 224-247.

Murphy, K. R., & Davidshofer, C. O. (1991). *Psychological testing: Principles and applications* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598-609.

Paulhus, D. L. (1990). *Assessing self-deception and impression management in self-reports: The balanced inventory of desirable responding--Reference manual for version 6.* Unpublished manual.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego: Academic Press.

Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches.* Dubuque, IA: Wm. C. Brown.

Rosenfeld, P., Doherty, L. M., & Carroll, L. (1987). Microcomputer-based organizational survey assessment: Applications to training. *Journal of Business and Psychology, 2,* 182-193.

Rosenfeld, P., Doherty, L. M., Vicino, S. M., Kantor, J., & Greaves, J. (1989). Attitude assessment in organizations: Testing three microcomputer-based survey systems. *Journal of General Psychology, 116,* 145-154.

Rosenfeld, P., Giacalone, R. A., Knouse, S. B., Doherty, L. M., Vicino, S. M., Kantor, J., & Greaves, J. (1991). Impression management, candor, and microcomputer-based organizational surveys: An individual differences approach. *Computers in Human Behavior, 7,* 23-32.

Rozensky, R. H., Honor, L. F., Rasinski, K., Tovian, S. M., & Herz, G. I. (1986). Paper-and-pencil versus computer-administered MMPIs: A comparison of patients' attitudes. *Computers in Human Behavior, 2,* 111-116.

Schuldberg, D. (1988). The MMPI is less sensitive to the automated testing format than it is to repeated testing: Item and scale effects. *Computers in Human Behavior, 4,* 285-298.

Skinner, H. A., & Allen, B. A. (1983). Does the computer make a difference? Computerized versus face-tc-face versus self-report assessment of alcohol, drug and tobacco use. *Journal of Consulting and Clinical Psychology, 51*, 267-275.

White, D. M., Clements, C. B., & Fowler, R. D. (1985). A comparison of computer administration with standard administration of the MMPI. *Computers in Human Behavior, 1*, 153-162.

Wilson, F. R., Genco, K. T., & Yager, G. G. (1985). Assessing the equivalence of paper-and-pencil vs. computerized tests: Demonstration of a promising methodology. *Computers in Human Behavior, 1*, 265-275.

Best Project Nomination Rationale

Originality/Technical Merit

Computerized surveys, as substitutes for paper-and-pencil surveys, are growing in popularity in military and civilian organizations as computer technology becomes widespread. The benefits offered by computerized surveys (e.g., elimination of missing responses, automated data entry) can be offset by respondents' concern that their anonymity may be compromised, affecting their responses and reducing validity. The empirical research literature on this question has been inconsistent: sometimes computerized surveys produced less valid responses than paper-and-pencil surveys, sometimes more, and sometimes they have not differed.

It is important to resolve these differences if computerized surveys are to be used in organizational management. The present research attempted to clarify the problem by replicating and extending a recent major study that questioned the validity of computerized, organizational surveys. Most studies used college students as subjects. The present research compared the responses of Navy recruits on either a computerized or a paper-and-pencil survey in an applied organizational setting.

Research Approach/Plan/Focus/Coordination

Navy recruits, split into appropriate groups, completed either a pencil-and-paper or computerized survey. Some responded anonymously, while others were self-identified. As expected, there was an increase in socially desirable responding when respondents were identifiable. However, scores for the computerized and pencil-and-paper conditions did not differ. Furthermore, the internal consistency reliabilities of the survey scales were similar for the two conditions.

Difficulty of Problem Addressed

The ambivalence of previous research findings posed a difficult problem for both researchers interested in survey methods and practitioners who administer organizational surveys. The present study utilized an applied organizational setting while maintaining strict experimental control. It is difficult to maintain

such control in applied settings. Thus, a direct test of the comparability of computer-based and traditional methods was achieved. The experimental rigor allowed for the conclusion that the methods did not differ significantly.

Potential Impact on Navy/Center Needs

The Navy frequently uses organizational surveys to assess perceptions, attitudes, and behavior tendencies of personnel. As computers in the workplace become more commonplace, the administration of large-scale computerized surveys becomes feasible. Information regarding the techniques needed to obtain valid results is needed before the technology can be implemented. The results of this exploratory research provide information favorable to the use of computerized surveys. They challenge claims that the quality of data obtained by computer is less valid than that obtained by traditional pencil-and-paper surveys.

Productivity

The effort was highly productive. The research was completed quickly. Several papers have been completed and submitted for publication. In addition, the study can contribute directly to productivity in conducting Navy surveys. Computerized surveys can be more accurate, reduce labor in entering data, and provide information to Navy managers more quickly. Developing the appropriate techniques for computerizing surveys can have important practical consequences.

Appropriateness of IED Support

The availability of computer technology to help in gathering important information on Navy personnel can bring about large increases in efficiency. However, we must be sure that the information obtained is a valid representation of personnel responses (i.e., that computerized method does not differ from the non-computerized method currently used). To help in this process, this study developed computer-based survey techniques and explored whether they obtained information comparable with standard survey methods. Such information has important implications for practice in determining whether we can capture the efficiencies inherent in computerized surveying.

Publication Status

Booth-Kewley, S., Edwards, J. E., & Rosenfeld, P. (in press). Impression management, social desirability, and computer administration of attitude questionnaires: Does the computer make a difference? *Journal of Applied Psychology.*

Booth-Kewley, S., Rosenfeld, P., & Edwards, J. E. (in press). Impression management and self-deceptive enhancement among Hispanic and non-Hispanic white Navy recruits. *Journal of Social Psychology.*

Edwards, J. E., Booth-Kewley, S., & Rosenfeld, P. (1991) *Computer administration of questionnaires: Impression management and social desirability.* Paper presented at the 33rd annual conference of the Military Testing Association, San Antonio, TX.

Independent Research

Progress Reports

The Role of Feedback in Computer-based Training

Michael Cowen

Abstract

Current research in computer-based training (CBT) provides little guidance as to when feedback should be provided and how to design feedback content. An experimental CBT lesson on how to operate a military phone system was administered to 80 Navy students. Results showed those who received delayed feedback significantly outperformed those who received immediate feedback. However, the delayed feedback subjects also spent significantly more time in the practice segment of the lesson. Moreover, only short term retention (i.e., immediately after the practice) of the material was measured. The follow-on work is designed to replicate the outcomes controlling for practice time and to test for long-term effects.

Introduction

Navy personnel often have difficulty operating the state-of-the-art programmable equipment employed in radar systems, communication systems, and weapon systems. These types of devices tend to be designed without adequate consideration of the user interface. Indeed, conceptual models of how the device works are used in the engineering of the device but engineers often give little thought on how the user will make it work (e.g., pushing buttons, flipping switches). Computer-based training systems have been developed to help users overcome the learning difficulties associated with operating these types of digitally controlled devices. Computers are used to teach device operation because they provide a safe environment for users to learn about how to operate equipment without endangering themselves and others or harming the equipment.

Feedback to student responses is an important design feature of CBT. Feedback provides information to the student about the correctness of the student's knowledge of the device procedures. Feedback is implemented unsystematically in current CBT systems. Although meta-analyses have demonstrated that computerized instruction can improve student achievement, there is no clear indication from the literature as to when and how much feedback should be provided during CBT. The

objective of this effort is to determine how manipulating the timing and the content of feedback in CBT affects learning to operate a digitally controlled device.

Two theories of cognition and skill acquisition, ACT* theory and instructionless learning have different implications for when and how much feedback should be provided during CBT. ACT* (adaptive control of thought) is a theory that describes the learning of procedures. ACT* is a theory of cognition and skill acquisition based on production systems. A production system is a hierarchical set of mental tasks consisting of condition-action pairs called productions. The condition-action pairs represent specific mental and physical actions that should occur if a particular state occurs in working memory. Implications of ACT* are that feedback that provides the correct response to a student error should be more effective than feedback in the form a "you are wrong" statement, and immediate feedback should be more effective than delayed feedback.

Instructionless learning, a type of discovery learning, suggests that users can learn how to work a digitally controlled device without the benefit of any written or verbal instructions. Users figure out how to work a device by discovering device behaviors as a result of user actions. During instructionless learning, the students form hypotheses about the syntax and semantics of the device switches, buttons, and dials. Implications of instructionless learning are that feedback in the form of a "you are wrong" statement as a response to a student error should be more effective than feedback that provides the correct response, and delayed feedback should be more effective than immediate feedback.

During the base effort, experimental computer-based training was administered to 80 Navy students. Instruction on how to operate a digitally controlled device was presented individually on an Apple Macintosh microcomputer using "HyperCard"-type software. This lesson used a modified drill and practice instructional strategy. The lesson had three parts: introduction, drill, and performance test. The introduction presented some general information about the digitally controlled device. It listed the features of the device and presented the name and location of each of the device's buttons. It did not present any information on how to work the device.

During the drill, the subjects received training on eight device tasks. Drill consisted of practicing each task by "clicking" with a mouse on a computerized graphic representation of the device.

The graphic representation had 16 active buttons. If an error was made, the CBT system provided feedback. There was four feedback conditions: immediate confirmatory feedback, immediate corrective feedback, delayed confirmatory feedback, and delayed corrective feedback.

Immediate feedback was defined as feedback provided the instant an error was made by the subject. In the immediate confirmatory feedback condition, feedback was a "wrong" indication by the computer to an error. In the immediate corrective feedback condition, feedback was a presentation of a single correct operation to an error. Also, in the immediate feedback conditions the subjects could not move to the next step until the correct step had been clicked. After all the correct steps had been entered for that task, the subject proceeded to the next task (i.e., activating another feature).

Delayed feedback was defined as feedback provided at the end of button-pushing for that task. End of button-pushing was specified as a click on a "done" button that was located on the graphic representation. The subject clicked the "done" button only after he had clicked a sequence of button-pushes that might have activated the target task. The subjects could not move to the next task until the correct sequence of steps has been entered. In the delayed corrective feedback condition, feedback was a presentation of the entire correct sequence of steps for that task as a response to any error in the entered sequence. In the delayed confirmatory feedback condition, feedback to student errors was a listing by the computer of the entered steps that were correct in terms of whether that keypress was a component of that task and whether that keypress was entered at the correct step.

During the performance test, the subject was asked to perform each of the device tasks on the graphic representation of the digitally controlled device. For each item, the subjects was required to click a sequence of button-pushes and then click on the "done" button. The items of the performance test was different than the items practiced during the drill. The number of items correct and the subject's time on each item was recorded. Control subjects were also tested. The control group was administered the introduction and performance test portions of the lesson, but not the drill.

Results from the base effort showed that those who received delayed feedback significantly outperformed those who received immediate feedback. However, the delayed feedback subjects also spent significantly more time in the practice segment of the

lesson. Indeed, increased practice time may be a desired characteristic in the design of CBT. However, an experiment in which the subjects in all the conditions have equal practice time may provide a more appropriate test of feedback in CBT. Moreover, only short-term retention (i.e., immediately after the practice) of the material was measured. It is unclear whether there was any long-term effects for delayed feedback. The follow-on work is an attempt to replicate the outcomes found in the base effort controlling for possible confounding variables.

Progress and Plans

Segments of the computer-based training developed for the base effort were revised so that the practice time would be increased during the CBT with immediate feedback and decreased during the CBT with delayed feedback. Practice time during CBT with immediate feedback was increased by requiring the subjects in this condition to practice each task twice.

Practice time during CBT with delayed feedback was decreased by showing these subjects how to do a task before they practice a task. Other features of CBT with delayed feedback were also modified in order to decrease practice time. These included not forcing subjects to start at the beginning of the task each time they wanted to redo a step (an "erase last entry" feature was implemented) and the use of only corrective feedback. In the base effort, it was found that the delayed condition with confirmatory feedback used significantly more practice time than the delayed condition with corrective feedback. It should be noted that these two groups performed about the same on the post-test.

In total, the follow-on study tested five conditions: three immediate corrective feedback groups and two delayed corrective feedback groups. The immediate feedback conditions consisted of (1) a condition identical to the immediate corrective feedback treatment used in the base effort, (2) a condition identical to the first using two practice trials, and (3) a condition identical to the first with the addition of having these subjects shown how to do a task before they practice a task. The delayed feedback conditions consisted of (4) a condition identical to the delayed corrective feedback treatment used in the base effort with the addition of the "erase last entry" feature, and (5) a condition identical to the fourth with the addition of having these subjects shown how to do a task before they practice a task.

Also, the follow-on study added a second administration of the post-test 24 hours after the experimental CBT.

The revised experimental CBT has been administered to approximately 150 Navy students awaiting instruction at sonar technician "A" school. Data analysis is underway. Results will be published in an NPRDC technical report.

Expected Benefit

In addition to substantially incrementing our knowledge of how people learn to operate digitally controlled devices, this effort made a considerable step toward designing a practical system for training. For a digitally controlled device of reasonable complexity, a low-cost simulation was implemented. Although some feedback groups performed better than others, all feedback groups significantly outperformed the control group. Therefore, with appropriate attention to implementation, this computer-based instructional system could impact the performance of a Navy system.

Publications and Presentations

Cowen, M. B. (June, 1991). *Computer-based training (CBT) for operating a digitally controlled device: The role of feedback.* Paper presented at the fourth Navy IR/IED symposium, Laurel, MD.

Cowen, M. B. (July, 1991). *The development of a computer-based instructional simulation for operating complex electronic products.* Paper presented at the fifth annual adult literacy and technology conference, Costa Mesa, CA.

Cowen, M. B. (October, 1991). *A comparison of four types of feedback during computer-based training (CBT)* (NPRDC-TR-92-2). San Diego: Navy Personnel Research and Development Center.

Experientially-based Learning of Multiple Roles

Robert F. Morrison

Abstract

The primary source of learning a job is the job itself, but little is known about the factors that enhance or inhibit such experience-based learning. Preliminary research has demonstrated that experience-based learning of a new, middle-management position is influenced by the characteristics of the individual, job, immediate work context, and more macro work and non-work environment. While these factors have been shown to affect the total position, it would be anticipated that different ones may be salient in the learning of some job roles but not others. The objective of this research is to contrast the predictor sets of variables for each of 11 middle-management job roles. The results will be used to supplement the model of experience-based learning (EBL) reported previously. The EBL model should aid an organization that rapidly and continuously transfers personnel among jobs, to establish policies and practices that optimize the learning/adaptability of the personnel being transferred.

Background and Problems

Preliminary research on managerial positions challenges the assumption that all effective learning must occur in a structured (classroom-like) environment. Such research indicates that the majority of learning occurs as a result of work experience (Brousseau, 1984; Fleischman & Mumford, 1989; Hall, 1991; McCall, Lombardo, & Morrison 1988; Morgan, Hall & Martier, 1979; Vineberg & Taylor, 1972). While a model and propositions covering an entire career have been proposed (Morrison & Hock, 1986), the attributes of its components were not defined in detail. This definition is imperative to the adequate explication of the career development process.

The speed with which experienced personnel (surface warfare department heads) learn a totally new position during an assignment appears to be influenced by a multitude of factors (Morrison & Brantner, in press). These factors include individual-differences, job-characteristics, immediate work context, and more macro work and non-work environmental

effects. Further, any position consists of numerous roles that may be learned at different rates and may be affected by different factors. It would be very useful in the development of theory as well as application to establish if the factors that influence the learning of different roles are significantly different.

A plethora of demands (e.g., joint service and materiel professional development) has forced policymakers to shorten billet and command tours until they are frequently less than 18 months. Such policies have been designed using manpower flow models without considering their effect on the officers' performance and career development. The fleet's personnel readiness and the effectiveness with which support activities perform are affected directly by the opportunity that officers have to develop the capability to learn the requisite knowledge and skill of each billet and to develop them beyond the level of mastery. Tour lengths that are too short do not provide the opportunity to develop, while ones that are too long make inefficient use of the officer force and may lower the officers motivation to perform at a high level or learn new tasks/jobs.

Objective

The broad objective of this research is to develop a generic model describing the factors that influence how long it takes an individual to develop an expert-level of skill in performing work. The specific objective is to develop, test, and modify a preliminary model of the learning that occurs while the incumbent is in a mid-level leadership position on a surface ship.

General Approach

A literature search was used to identify: (1) the steps that an individual goes through in learning how to perform a job to the point of mastery, (2) the parameters that contribute to the level of performance at entry, and (3) the factors influencing what is learned and how quickly it is learned. This information was used to form an initial model of the experiential learning process.

The preliminary model was revised using information acquired from interviews (Morrison & Brantner, 1989). Research (Morrison & Brantner, in press) was conducted to test a situationally specific model of experiential learning for the aggregate position of department heads on surface warfare ships.

With the support of the Office of the Chief of Naval Operations (OP-13) and Surface Forces, Atlantic Fleet and Pacific Fleet, data were obtained from 396 department heads representing 322 surface ships. These data emphasized perceptions of individual differences, job characteristics, context, and environment factors plus assessments of the extent that the officers had learned each of the 11 roles required to do their jobs. The initial analyses were done using a sample of 324 officers in which their CO's had reported the officers' positions on the learning curve. Since the COs' evaluations were too skewed to use, it was decided to use only the officers' data.

Progress

Additional data were loaded into the database enlarged the total sample size from 324 to 396. For the purposes of this research, the usable sample size increased from 295 to 365. This improved the stability of the learning curves reported by the officers in both the first and second half of the split tour. The number of officers in the single, long tour increased marginally from 40 to 42, and there was no effective change in the stability of that group's learning curves.

Intercorrelations were high among the officers' self-ratings of their positions on the learning curves for the 11 roles required in the performance of the department head position. Such correlations were high enough that the ratings could be combined additively to represent a position on the learning curve for the over-all position. However, the shapes and/or levels of the learning curves are nearly all significantly different from each other. In addition, the predictors of position on each of the 11 learning curves vary among the roles as noted below.

Independent path analyses were conducted on each of the 11 roles. In each instance, only the individual differences, job characteristics, context, and environmental factors that made a direct, significant contribution to explaining the variance in position on the learning curve beyond time on the job were retained in each model (see Table 1). While the second-level factors that predicted the first-level factors described in the previous sentence were investigated, only the first-level factors are summarized in this report.

Table 1

Predictors of Position on the Learning for Various Roles

Variable	Description
Mo	Number of months on the job.
Individual Differences	
SE	Self-efficacy (How capable of doing the job based on past experiences and ability).
SQT	Time to pass professional qualification exam.
TQ	Total professional qualifications earned.
Job Characteristics	
JCB	Lack of job challenges (Job is boring).
JCD	Job is demanding (i.e., difficult and under pressure).
RCL	Role clarity (How clearly expressed are the command goals
RCO	Role Complexity: Technology of equipment.
ROW	Role Overload: Hours spent with work group last week.
Context	
SQ	Subordinate Quality (Enlisted personnel are good performers).
OMA	Organization Mission: Amphibious.
OCN	Entered organization while in operational stage of cycle.
OCC	Currently in maintenance stage of organization's cycle.
LC	Leadership climate.
Environment	
Mar	Marital status affects learning.
Num	Number of Dependents.

The significant first-level factors that predicted position on the learning curves beyond time on the job are noted for six of the 11 department head job roles (note the increase in R^2 between the base equation for months on the job and the full equations).

Working Relationship Roles

- Immediate Subordinates $= 1.56\,Mo - 2.12\,Mo^2 + .96\,Mo^3$ $(R^2 = 16)$

    Immediate Subordinates $= 1.76\,Mo - 2.77\,Mo^2 + 1.30\,Mo^3 + 11\,SE + .21\,JCB + .12\,RCL + .17\,SQ + .14\,OMA + .11\,OCN - .10\,Mar$ $(R^2 = .28)$

- Distal Subordinates $= 1.82\,Mo - 2.56\,Mo^2 + 1.09\,Mo^3$ $(R^2 = .16)$

    Distal Subordinates $= 1.98\,Mo - 3.00\,Mo^2 + 1.30\,Mo^3 + .12\,SE + .19\,JCB + .20\,SQ + .14\,OCN - .13\,Mar$ $(R^2 = .28)$

- Peers $= 1.55\,Mo - 2.29\,Mo^2 + 1.12\,Mo^3$ $(R^2 = .13)$

    Peers $= 1.93\,Mo - 3.26\,Mo^2 + 1.69\,Mo^3 + .27\,JCB - .13\,JCD + .12\,ROW + .26\,LC$ $(R^2 = .31)$

- Superior $= .88\,Mo - .63\,Mo^2$ $(R^2 = .11)$

    Superior $= .88\,Mo - .65\,Mo^2 + .14\,SE + .19\,JCB - .10\,Mar$ $(R^2 = .35)$

Technical Roles

- Material Condition $= .84\,Mo - .48\,Mo^2$ $(R^2 = .17)$

    Material Condition $= .85\,Mo - .54\,Mo^2 + .13\,SE - .10\,SQT + .15\,JCB + .12\,RCL + .13\,RCO - .10\,OCC - .16\,Mar + .10\,Num$ $(R^2 = .30)$

- Technical Knowledge $= .75\,Mo - .36\,Mo^2$ $(R^2 = .18)$

    Technical Knowledge $= .84\,Mo - .47\,Mo^2 + .20\,SE + .17\,TQ + .16\,JCB + .13\,RCO$ $(R^2 = .30)$

These two classes of roles provide an interesting contrast since cognitive concerns such as the attainment of professional qualifications and the technology of equipment are contributors to learning technical roles but not working relationship roles. The context of the job appears to be more important in learning working relationship roles, especially those with subordinates, than technical roles. Self-efficacy, job challenges, and marital support are consistent predictors of enhanced learning roles regardless of their social or cognitive content.

Plans

In FY92, confirmatory factor analysis and structural equation modeling using LISREL 7 will be applied to these data. They should either provide a verification of these results or question their validity.

A paper, (Morrison, in review) based on the path analyses from this project, has been submitted for presentation in a poster session at the annual meeting of the American Psychological Society.

Awards

Morrison, R. F., & Brantner, T. M. (1991). What affects how quickly a new job is learned? In J. L. Wall, & L. R. Jauch (Eds.), *Academy of management best paper proceedings 1991* (pp. 52-56). Monroe, LA: Academy of Management.

References

Brousseau, K. R. (1984). Job-person dynamics and career development. *Research in Personnel and Human Resources Management, 2,* 125-154.

Fleischman, E. A., & Mumford, M. D. (1989). Individual attributes and training performance. In I. L. Goldstein (Ed.), *Training and development in organizations* (pp.183-255). San Francisco: Jossey-Bass.

Hall, D. T. (1991). Twenty questions: Research needed to advance the field of careers. In R. F. Morrison & J. Adams (Eds.), *Contemporary issues in career development,* Hillsdale, NJ: Lawrence Erlbaum.

McCall, M. W., Lombardo, M ' ' , & Morrison, A. M. (1988). *The lessons of experience* .ington, MA: Lexington Books.

Morgan, M., Hall, D. T. & Martier, A. (1979). Career development strategies in industry: Where are we and where should we be? *Personnel, 56*(2), 13-31.

Morrison, R. F. (in process). Learning work roles: Working relationships vary from task knowledge.

Morrison, R. F., & Brantner, T. M. (1989). *An experience-based learning model: A pilot study* (unpublished manuscript). San Diego: Navy Personnel Research and Development Center.

Morrison, R. F., & Brantner, T. M. (in press). What enhances or inhibits learning a new job: A basic career issue. *Journal of Applied Psychology*.

Morrison, R. F., & Hoch, R. R. (1986). Career building: Learning from cumulative work experience. In D. T. Hall (Ed.), *Career development in organizations* (pp. 236-273). San Francisco, CA.

Vineberg, R., & Taylor, E.N. (1972). *Performance in four Army jobs by men at different aptitude (AFQT) levels: 3. The relationship of AFQT and job experience to job performance* (HumRRO Tech. Rep. 72- 22). Alexandria, VA: Human Resources Research Organization.

Artificial Neural Networks and Training

Jan Dickieson and Lew Gollub

Abstract

The development of effective training methods requires predicting how the trainee will respond to the training procedure. At present, only qualitative models and intuitions can guide the training developer. A quantitative model of human behavior, as it changes through training, would facilitate the development of optimal training methods and materials by providing a platform for rapidly pretesting procedures prior to a more costly field test. This project used neural network analysis techniques to develop a model of the acquisition of some aspects of a Navy training task (Air Intercept Controller). This model will then serve as the basis for predicting the effects of changes in training conditions.

Rationale

The primary goal of this research project has been to develop an artificial neural network (ANN) model of learning and performance on a Navy relevant task. Such a model can then be used to predict the effects of changes in training conditions with the long-term goal of developing a model that can predict the effects of novel combinations of training parameters. This research will attempt to develop a neural network model of a learner's response to complex, demanding environmental situations, such as Air Intercept Control (AIC), Air Traffic Control, or Air Combat Maneuvering (ACM).

Artificial neural networks are a recently developed method for modeling complex systems. A great variety of processes have already been modeled with this approach, including perceptual systems for detecting SONAR profiles, entrance criteria for United States Naval Academy (USNA) applicants, and financial characteristics of good and bad loan applicants as well as numerous physiological and behavioral conditioning systems. The diversity of applications indicates the flexibility of the approach.

In essence, an ANN is a collection of highly interconnected simple processing elements (PEs) analogous to the neurons of the nervous system. Each PE receives one or more inputs, which

may "stimulate" or "inhibit" it. As a result of these inputs, the PE may produce an output signal that is, itself, an input to another PE or the system output. An ANN model consists of a specific structure of interconnected elements and rules governing how the model changes. The influence of input information (similar to sensory neurons) on interconnecting ("hidden") elements and on output ("motor") elements is adjusted according to learning rules.

Although there are a number of specific findings that describe the acquisition of high-performance skills, no successful unified model of learning has been developed for this important class of activities. Since these situations involve changes in response with repeated presentation of learning scenarios, ANNs offer a plausible modeling approach, since they permit a broad range of model structures and learning rules.

Relatively few studies have been reported in which ANNs have been used to model human or other complex behavior. However, successful attempts have modeled ACM decision making as well as simple conditioning and discrimination learning in laboratory animals. Thus, it is plausible to apply ANNs to the training environment.

Approach

The AIC task was chosen for study. A trainee examines the Navy Tactical Data System (NTDS) display screen, fixing a target aircraft by moving a computer trackball, and issuing commands to a controlled aircraft to intercept (or avoid) the target. This task involves a series of perceptual and fine motor response requirements as well as complex decision-making. This task thus demands a variety of skills whose acquisition can be studied under different training conditions.

The research plan involves three phases: (1) development of a simulation of the AIC task on personal computers (PCs), so that studies can be performed on a variety of computer equipment, (2) collection of training data on simple tasks in which stimulus and response conditions can be recorded in detail, and (3) selecting and developing a model involving stimulus and time (training) inputs for subject behaviors.

Progress

Software was obtained to run the task from Dr. Schneider of the University of Pittsburgh. The original program written in Pascal was recoded in C-language. Additionally, several training lessons were written involving low-level intercept skills: hooking and intercept path estimation. The recording of trainee data was improved, so that we have a complete record showing the state of the display, the subject's response, feedback messages, and time of occurrence.

We continued to review the published literature for applications of ANNs to learning and performance. Because ANNs have not been applied to skill learning situations, a thorough review of related studies on simple classical and operant conditioning was undertaken to examine points of similarity and difference from skill learning.

Reductions in FY92 project funding after the first quarter resulted in suspending the research until funds are available.

An Exploratory Examination of Artificial Neural Networks as an Alternative to Linear Regression

Jan Dickieson and Charles Wilkins

Abstract

Accurate prediction of the future performance of personnel is a vital piece of information, which the Navy must use to make appropriate decisions. These predictions are most commonly made using some type of linear-regression-based method. This means that not all available information (e.g., nonlinear relationships) are used in making predictions. Artificial Neural Network (ANN) technology, on the other hand, allows prediction models to be created, which do take into account nonlinear relationships. This study was designed primarily to explore the feasibility of applying neural network technology to a typical Navy problem, specifically, predicting premature attrition from the U.S. Naval Academy. Attrition data from various classes of the Naval Academy was used to predict premature attrition using both standard linear regression and a variety of ANN models. Under robust conditions, it was found that the predictive efficacy of the ANN models was superior to that of linear regression, leading to the conclusion that ANN models do indeed merit additional research. Future research into how to take full advantage of ANN technology is discussed.

Background and Problem

The Navy makes important decisions everyday about a variety of personnel issues based, in part, upon the prediction of future performance. To the degree that these predictions are less than optimal, the resulting decisions may have an adverse impact. The most common technique used to predict performance is simple multiple linear regression. However, any nonlinear relationships that exist in the data being studied cannot be accounted for by simple linear regression-based procedures. Nonlinear regression techniques exist; however, these typically require a priori specification of a model of the nonlinear relationship. Unless the form of the nonlinearity is well understood, it is difficult to choose such a nonlinear model.

In recent years, new techniques have been developed in which computers 'learn' the relationships between a set of variables,

without needing a priori information about the relationships. These techniques, called ANNs, have been applied to a wide variety of problems which include controlling robotic arm movement (Martinetz & Schulten, 1990), modelling the spelling ability of brain-damaged persons (Olson & Caramazza, 1988) and artificial intelligence problems such as learning to play backgammon (Tesauro, 1990).

Far less research has been conducted on the efficacy of using neural networks for prediction. Recently, a number of researchers have proven that neural networks (at least theoretically) serve as universal approximators under a wide variety of conditions (Hecht-Nielsen, 1990; Hornik, Stinchcombe & White, 1990). However, neural networks are iterative techniques, and many questions remain, such as how to best choose the neural network configuration to use, deciding when to terminate the iteration process, how to best encode the input data, etc. Therefore, in practice at least, there remain many unresolved questions about the predictive efficacy of neural networks.

This current research was designed to examine some of these practical issues in the context of a serious Navy problem, namely premature voluntary attrition from the United States Naval Academy. The Academy is one of the Navy's most important resources for recruiting and training top quality officers. It is an expensive resource, however, with a 4-year course of study costing approximately $153,000 per person (GAO, 1991). When a midshipman prematurely attrites from the Academy, the money that has been invested in his or her training has been essentially wasted. In light of reduced operating budgets, it is important to study ways in which attrition can be reduced, since even a small reduction could lead to substantial savings.

Objective

The objective of this work was to explore, empirically, the advantages and drawbacks of using ANNs as an alternative to linear regression techniques for various prediction problems of concern to the Navy, specifically predicting premature voluntary attrition from the Naval Academy. The study was designed to determine whether ANNs showed sufficient promise as a prediction methodology to warrant additional research into how best to apply ANNs to Navy personnel and manpower problems.

Approach

The basic approach of this study was to develop prediction models using both linear regression and ANNs and to then compare the predictive efficacy of the two methods. The study used data from three recent classes of the Naval Academy. The three classes in this study will be referred to as Class I, Class II, and Class III.

The prediction models for both methods were developed using the seven predictors currently used by the Academy to evaluate candidates. These variables are SAT-Verbal, SAT-Quantitative, high school rank in class, recommendations from high school officials, extracurricular activity score, technical interest score, and career interest score (Wahrenbrock & Neumann, 1989). These seven predictors are currently used because they have the strongest linear relationship with criteria of interest to the Academy (academic performance, military performance, choice of major, and the criterion of interest in this study--voluntary attrition). Choosing these predictors does not take full advantage of the ANNs ability to model nonlinear relationships, but it is an appropriate initial comparison, since these are the predictors currently used by the Naval Academy.

The first part of this study consisted of using Class I to calculate an appropriate regression equation, utilizing the same stepwise procedure currently used by the Naval Academy. This regression equation was then cross-validated using data from Class III, which was completely independent of Class I.

The procedure for the ANN portion of the study was more complicated, because ANNs are trained with an iterative procedure, and there is no simple stopping criterion for ANNs with the type of data used in this study. Therefore, a **two-phase cross-validation paradigm** was developed. Six different neural networks of varying characteristics were used (see Table 1). Each ANN was trained on data from Class I in increments of 10,000 iterations. The ANNs were then cross-validated on Class II at each number of iterations. These results were used to determine the number of iterations for training, and then the network was cross-validated again on Class III, which was independent of both Classes I and II.

Two different stopping criteria were examined in this study. Criterion A was simply the number of iterations that provided the maximum cross-validation correlation coefficient for Class II. Criterion B was the midpoint of the range of iterations for

Table 1

Neural Network's Characteristics

Network	Architecture	Inputs	Hidden Layer 1	Hidden Layer 2	Outputs
1	Backpropagation	7	14	0	1
2	Backpropagation	7	7	0	1
3	Functional Link	7	7	0	1
4	Functional Link	7	4	3	1
5	Backpropagation	7	21	0	1
6	Backpropagation	7	2	0	1

which ANN provided a higher cross-validation correlation coefficient than linear regression. This yielded a rougher estimate of the optimal stopping criterion, but one which was hopefully more robust.

Results

Stepwise multiple linear regression was conducted on the data in Class I using SPSSX. The resultant regression equation was then used to predict attrition for Class III. The correlation between predicted attrition and actual attrition was then calculated for Class III. Attrition is very difficult to predict, and because it is a dichotomous variable, with only a small percentage of people actually attriting, the size of the correlation coefficients are diminished. The correlation in this case was found to be .0561. This served as a baseline for comparisons with ANNs.

The ANNs were trained and tested in increments of 10,000, up to 200,000 iterations. For each increment, the correlation between predicted attrition and actual attrition was calculated. There was a broad range of stopping values for which the ANNs had higher correlations than linear regression did. These are shown in Figure 1.

The two-phase cross-validation paradigm was used to observe how well information about Class II could be used to choose a stopping criterion for Class III. Criteria A and B were calculated for each of the six networks, and using these stopping criteria, the cross-validated correlations were calculated for Class III. For all six networks, both Criteria A and B yielded correlations higher than those provided by linear regression (see Table 2).

An Exploratory Examination of ANNs as an Alternative to Linear Regression
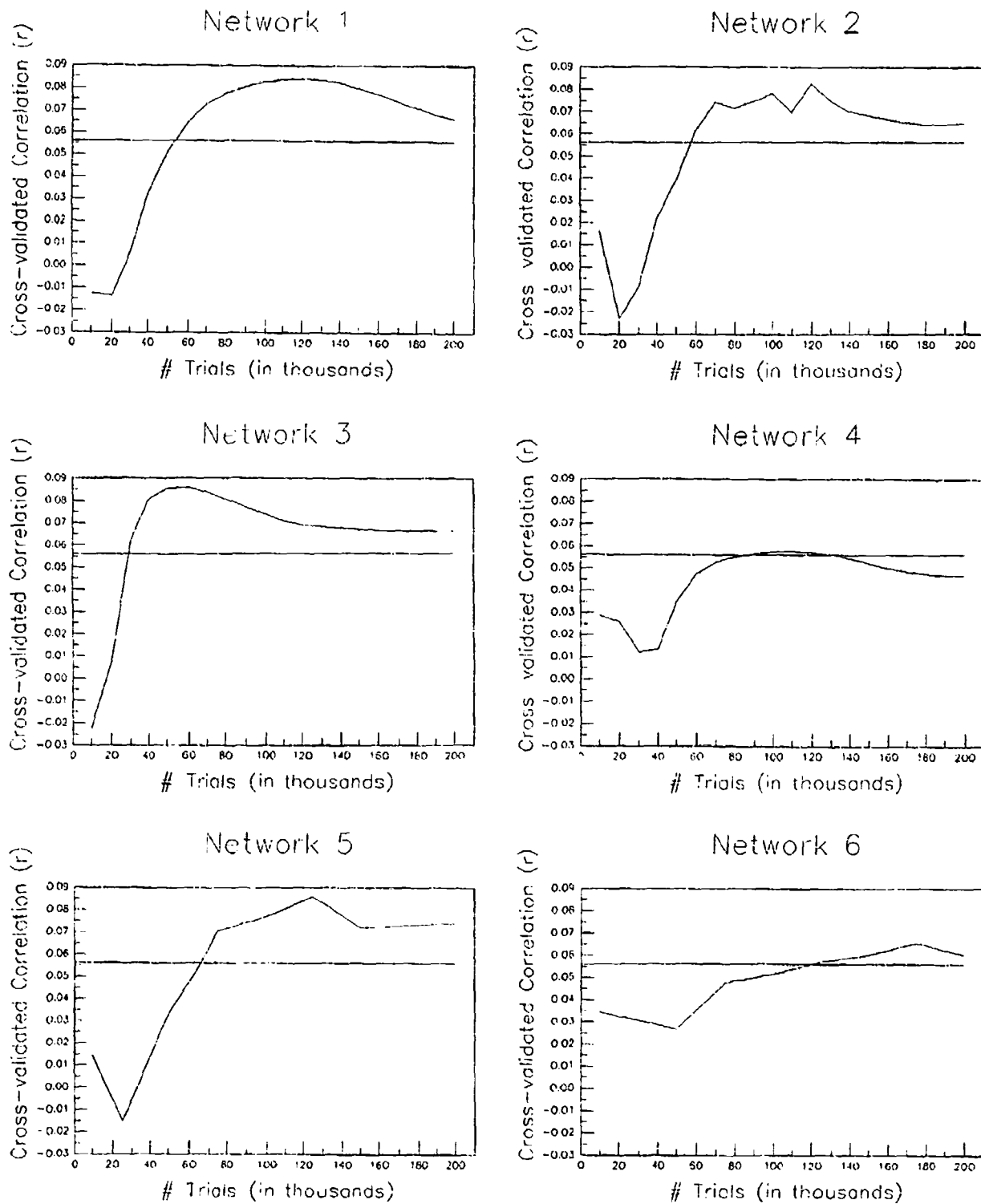
Note. The horizontal line represents the correlation for linear regression.

**Figure 1. Cross-validated correlation coefficient for Class III
for varying stopping criteria.**

Table 2

Class III Cross-validated Correlation Coefficients

Network	Regression	NN-Criterion A	NN-Criterion B
1	.0561	.0846	.0806
2	.0561	.0806	.0762
3	.0561	.0854	.0858
4	.0561	.0577	.0577
5	.0561	.0860	.0769
6	.0561	.0657	.0657

Discussion and Conclusion

The results of this study indicate that ANNs show great promise as an alternative method to linear regression for prediction. These results demonstrate that, if properly used, ANNs can have higher predictive efficacy than linear regression. This is encouraging since ANNs are not limited to linear relationships between the input and output data, as linear regression is.

The main drawback to ANNs are their iterative nature. They are more difficult to use than linear regression, and, if trained for too many or too few iterations, they will lead to worse performance than linear regression. Still, this study certainly indicates that further research is warranted.

One of the primary areas of research planned for the future is an examination of other predictor variables. Predictors that have strong, but nonlinear relationships with attrition would improve the ability of ANNs to predict attrition, although they would not be very helpful to linear regression. Also, only a few different networks were included in this study. Further research on how to best choose the configuration of ANNs would be very useful. Research in these areas is currently underway at NPRDC.

In summary, preliminary research has shown neural networks to be a potentially powerful method for improving prediction of a number of variables of importance to the Navy.

References

General Accounting Office (GAO). (1991). *DOD service academies: Improved cost and performance monitoring needed* (GAO/NSIAD- 91-79). Washington, DC. Author.

Hecht-Nielsen, R. (1990). *Neurocomputing*. Addison-Wesley, Reading, MA.

Hornik, K., Stinchcombe, M., & White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks, 3*, 551-560.

Martinetz, T., & Schulten, K. (1990). Hierarchical neural net for learning control of a robot's arm and gripper. *Proceedings of IJCNN-90, San Diego, 2*, 747-752.

Olson, A. C., & Caramazza, A. (1988). *Lesioning a connectionist model of spelling*. Paper presented at Venice III: Conference on Cognitive Neuropsychology and Connectionism, Venice, Italy.

Rumelhart, D., & McClelland, J. (1986). *Parallel distributed processing: Microstructure of cognition: Vol. 2. Psychological and biological models*. Cambridge, MA: MIT Press.

Tesauro, G. (1990). Neurogammon: A neural-network backgammon program. *Proceedings of IJCNN-90, San Diego, 3*, 33-39.

Wahrenbrock, A. L., & Neumann, I. (1989). *United States Naval Academy longitudinal officer database documentation*. San Diego: Navy Personnel Research and Development Center.

A Comparison of Artificial Neural Networks and Linear Regression for Dichotomous Criterion Prediction

William A. Sands and Charles A. Wilkins

Abstract

Many important criteria in military personnel research are dichotomous or dichotomized (e.g., successful completion of first-term obligated service vs. premature attrition). Frequently, prediction of dichotomous criteria is done using Ordinary Least-Squares Linear Regression (OLS-LR) techniques. This study was designed to develop, evaluate, and compare alternative prediction models for forecasting dichotomized criteria, using Artificial Neural Network Back-Propagation (ANN-BP) technology. Computer-simulated data distributions were created and used to evaluate the cross-validated predictive efficacy of OLS-LR and ANN-BP under a variety of personnel selection decision/outcome situations. Classification accuracy, defined as the proportion of correct selection/rejection decisions, was the basis for comparing the two different approaches. Based upon the highly favorable results obtained with the ANN-BP model, the authors concluded that this approach to predicting dichotomous criteria appeared quite promising and that further research was definitely merited.

Background and Problem

Many of the important criteria for military personnel prediction problems are dichotomous (or continuous variables that have been dichotomized). An excellent example of these dichotomous criteria is successful completion of first tour of obligated service vs. premature attrition. The efficacy of prediction models for forecasting these dichotomous criteria is a paramount issue in personnel selection research.

Purpose

The purpose of this research was to compare and contrast the predictive utility of the OLS-LR model with an ANN-BP approach for personnel selection. Use of ANN models for a variety of research problems has been receiving increasing attention (Hecht-Nielsen, 1990; Khanna, 1990; McClelland &

Rumelhart, 1986; NeuralWare, 1991a; Rumelhart & McClelland, 1986). Ideally, this comparison should provide results that will allow generalization to a diverse set of personnel selection situations. This goal suggested the use of computer-simulated data. These data have the advantage of being "well-behaved" (in a statistical sense). Use of any real empirical dataset runs the risk of limiting (perhaps severely) the extent to which the results may be generalized.

Approach

The ANN technique used in this study was "back-propagation," as implemented by the NeuralWorks Professional II/PLUS software (NeuralWare, 1991b) on a Macintosh IIfx microcomputer. The dimensions of the study included the following:

1. Function forms of the data distributions.

2. Total sample sizes.

3. Errors--the degree to which the data points in the distribution deviate randomly from the ideal function form. This is measured by the standard deviation of the points around the function form. In the linear case, this error can be transformed to a validity coefficient (the correlation between the predictor and criterion).

4. Base rates--the proportion of successes before introducing a new selection instrument.

5. Selection ratios--the proportion of persons selected, using the new selection instrument.

6. Sample splits--the allocation of simulated persons ("simulees") in each total sample into two subsets: the development sample (used to develop a prediction model) and the evaluation sample (used to evaluate the prediction model).

The error of the distributions was chosen so as to yield desired validities in the linear case (.05, .25, .50, .75, and .90). The errors corresponding to these target correlations were used to generate total bivariate distributions for three sample sizes (100, 500, and 5,000) for each function form (linear and curvilinear). Then, the simulees in these total distributions were allocated to

development or evaluation samples, with the following
alternative splits (20-80%, 50-50%, & 60-40%).

The vector of scores was rank-ordered and then dichotomized
according to the desired base rate, producing two groups
(successes and failures). This procedure was followed separately
for each alternative base rate (.05, .25, .50, & .95).

An OLS-LR model was determined for each development
sample. These models were used to predict criterion scores for
each simulee in the associated evaluation sample. These
evaluation sample simulees were then rank-ordered by the
predicted criterion score. Alternative selection ratios (.05, .25,
.75, & .90) were imposed, dividing the evaluation sample into
those who would be selected and those who would be rejected at
each specified selection ratio.

At this point, the actual criterion status (success vs. failure) and
selection vs. rejection status for each simulee were available.
This allowed the formation of four decision-outcome
combinations: (1) correct acceptances, (2) erroneous
acceptances, (3) correct rejections, and (4) erroneous rejections.
This information was combined into the total number of correct
decisions and the total number of erroneous decisions. The
proportion of correct decisions ("hit rate") was employed as the
measure of effectiveness for comparing the OLS-LR approach to
the ANN-BP approach, under each combination of conditions
(function form, sample size, degree of error from the ideal
function form, base rate, selection ratio, & sample split).

Results

There were 62 comparisons in which there was a significant
difference between the two methods ($p < .001$). All of these
significant differences were observed for the curvilinear
distributions; none was observed for the linear distributions.
Sixty-one of these significant differences favored the ANN-BP
model over the OLS-LR model. Fifty-six of the 62 significant
differences were observed in the largest sample size (5,000), 6 in
samples of 500, and none when the sample size was 100. The
number of significant differences ($p < .001$) did not appear to be
related to base rates, selection ratios, or sample splits.

Discussion and Conclusion

The results of this study concerning the use of the ANN-BP approach for predicting dichotomous criteria are quite encouraging. One of the major advantages of the ANN-BP approach is that the researcher does not need to know the most appropriate function form for a dataset. In theory, the ANN-BP approach should seek out and discover the most effective prediction system. The OLS-LR model, on the other hand, will perform quite well on a dataset where the underlying relationship between the predictor and criterion is linear, but substantially less well when the underlying relationship is nonlinear. Unfortunately, the nature of the underlying relationship between a predictor variable and a criterion variable is frequently unknown. This study showed that a single ANN-BP configuration with a single fixed number of training iterations did as well as, or outperformed the OLS-LR approach under a wide variety of base rates, validities, selection ratios, and sample splits. The two factors that did have an important impact on producing significant differences between the two approaches were function form and sample size.

In the linear function form case, ANN-BP performed comparably to the OLS-LR approach. This finding is as much as one could hope, given that the OLS-LR will perform optimally for this function form. In the curvilinear case, the ANN-BP approach outperformed the OLS-LR approach in many instances. This is not surprising, since the OLS-LR procedure is trying to fit a linear model to curvilinear data. While it is true that an OLS-LR model could be specified to fit a certain nonlinear function, such a function would have to be specified in advance by the researcher. This is also true of traditional nonlinear regression approaches (e.g., logistic regression). One advantage of the ANN-BP approach is that no such prior functional form specification is required, and a single ANN-BP model can fit many diverse functional forms.

The ability of the ANN-BP approach to outperform the OLS-LR approach was related to sample size. When the size of the total sample (including the development and evaluation subsamples) was 100, no significant differences were observed ($p < .001$). When the total sample size was 500, ANN-BP outperformed the OLS-LR approach in six of 240 comparisons, while the remaining comparisons showed no significant differences ($p < .001$). When the total sample size was 5,000, there were 55 cases in which ANN-BP outperformed OLS-LR, 1 in which

OLS-LR was superior, and the remaining cases showed no significant differences ($p < .001$). In view of the fact that only one ANN-BP configuration and stopping criterion were employed for the large variety of conditions examined, these finding are quite encouraging. Clearly, the ANN-BP approach was quite robust, often producing superior performance. Examination of the one case wherein OLS-LR significantly outperformed ANN-BP ($p < .001$) suggested that the explanation involved the choice of a single ANN-BP configuration and stopping rule for all combinations of dimensions studied. Indeed, when the same network configuration was trained for an additional 100,000 iterations on the development sample data, the ANN-BP model significantly outperformed the OLS-LR model on the associated evaluation sample data.

In summary, a crude ANN-BP approach was found to perform as well as the OLS-LR approach under a wide variety of circumstances, and to significantly outperform the OLS-LR approach when sample sizes were large and the underlying function form deviated from a linear form. While additional research needs to be done to determine how to most effectively use ANN-BP models, it seems quite clear that this approach is a promising tool for military personnel research.

Papers and Presentations

Sands, W. A. (in press). Artificial networks: A tool for psychologists. *Proceedings of the 33rd Annual Conference of the Military Testing Association.* San Antonio, TX.

Sands, W. A., & Wilkins, C. A. (in press). Artificial neural networks for personnel selection. *Proceedings of the 33rd Annual Conference of the Military Testing Association.* San Antonio, TX.

References

Hecht-Nielsen, R. (1990). *Neurocomputing.* Reading, MA: Addison-Wesley Publishing Co.

Khanna, T. (1990). *Foundations of neural networks.* Reading, MA: Addison-Wesley Publishing Co.

McClelland, J. L., & Rumelhart, D. E. (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition-psychological and biological models* (Vol. 2). Cambridge, MA: The MIT Press.

NeuralWare (1991a). *Neural computing.* Pittsburgh, PA: NeuralWare, Inc.

NeuralWare (1991b). *Reference guide.* Pittsburgh, PA: NeuralWare, Inc.

Rumelhart, D. E., & McClelland, J. L. (Eds.) (1986). *Parallel distributed processing: Explorations in the microstructure of cognition-foundations* (Vol. 1). Cambridge, MA: The MIT Press.

Independent Exploratory Development

Progress Reports

A Large-scale Model to Maximize Navy Unit Personnel Readiness

Iosif A. Krass and Theodore J. Thompson

Abstract

The problem of optimally (re)allocating Navy personnel to combat units is compounded by several considerations: availability of trained personnel, staffing of positions by occupation groups or ranks, and maintaining an acceptable level of readiness. We formulate this problem as a network flow problem with side constraints. An additional, non-network variable measures the readiness level. The problem size can grow to as many as 36,000 arcs and 17,000 nodes with 3,700 side constraints. We develop two numerical methods for efficiently solving this problem. One method is based on a heuristic and is able to provide a feasible solution to the problem in reasonable time. We also develop an application of the Linear-Quadratic Penalty (LQP) method to exploit the embedded network structure by placing the side constraints into the objective function. The resulting nonlinear network program is solved using a simplicial decomposition of the network constraint set. Numerical results with both solution methods are provided.

Background

Every Navy combat unit is required to report the status of personnel readiness for the unit. The goal of the personnel unit readiness report is to ensure that combat ships and squadrons have sufficient personnel with specific skills to operate. Each combat unit is characterized by a number of functional mission areas such as mobility, anti-air warfare, or submarine warfare. A mission area within a unit requires personnel with different skills to support operational capabilities. Personnel skills are defined by rating or occupation and pay grade. A unit's capability to perform its functions in all its mission areas is referred to as "readiness."

The problem of maximizing personnel readiness of the U.S. Navy fleet can be formulated as a large scale, mixed integer problem. The size of the readiness problem is approximately 50,000 variables and 30,000 constraints. Commercial linear programming packages cannot solve this size problem, within reasonable time and space requirements, even using a super

computer. A possible exception is the AT&T KORBX machine, which is an integrated hardware and software optimization system. We developed a decomposition method, capable of solving this problem on moderate sized mainframe computers (e.g., NPRDC's IBM 4381).

The readiness problem can be formulated using existing data files. The demand side of the problem can be defined from manning files. These files provide the shortage or excess of personnel within pay grade, skill, and mission area for readiness activities. The supply side of the problem is defined from Enlisted Projection System files. These files contain projected number of people by skill, pay grade, and composite available to fill job openings. Composite defines sea or shore eligible.

Objective

The objectives of this work are (1) formulate an optimization model to maximize personnel unit readiness, and (2) develop a method capable of solving this model.

Approach

Our approach to this problem is:

1. Formulate the problem as a *network with side constraints*. Use maximal network and minimal side constraints in the formulation.

2. Attempt to solve the problem using NETSID.

3. Develop a heuristic solution technique.

4. Solve the problem using NETSID with the heuristic solution as a starting point.

5. Attempt to find an alternative solution technique if unsuccessful with NETSID.

Results

A mathematical model was formulated, which presents personnel unit readiness as an optimization on a network with side constraints. Based on this mathematical model, the problem

was decomposed into a large scale network (about 22,000 nodes and 50,000 arcs) with a relatively small number (about 4,000) of side constraints. The model generator was written in FORTRAN and debugged. We attempted to solve this problem with NETSID but were unsuccessful. NETSID is a state-of-the-art 'network with side constraints' code that we obtained from Professor Kennington of Southern Methodist University.

We then developed a heuristic algorithm to solve the readiness optimization problem on our IBM 4341. This heuristic solution was then used a the starting point for NETSID. This was also unsuccessful. However, we now have what we believe is a good heuristic algorithm for solving the problem.

Our work on this problem has generated interest in the academic community. Professor Stavros Zenios from Warton School, University of Pennsylvania said that he has a code that can solve this size problem. We provided the model specifications to him. They successfully solved the problem using the LQP algorithm. We are jointly preparing a paper (see below) for publication.

Papers and Presentations

Krass, I. A., Pinar, M. C., Thompson, T. J., & Zenios, S. A. (in process). *A Large-scale model to maximize Navy unit personnel readiness.*

Pinar, M. C., & Zenios, S. A. (1991). *A smooth penalty algorithm for network structured problems and its application to naval personnel assignment.* Paper presented at the Organizational Research Society of America/The Institute of Management Sciences Joint National Meeting, Anaheim, CA.

Krass, I. A., & Thompson, T. J. (1991). *Decomposition method for solving a large-scale model to maximize Navy unit personnel readiness.* Paper presented at the Organizational Research Society of America/The Institute of Management Sciences Joint National Meeting, Anaheim, CA.

Krass, I. A., & Thompson, T. J. (1991). *Navy Fleet Personnel Readiness Problem.* Paper presented at the Organizational Research Society of America/The Institute of Management Sciences Joint National Meeting, Nashville, TN.

Transitioning Research Knowledge into Practice: Information on What Works in Training to Aid Training Practitioners.

William E. Montague and Frederick G. Knirk

Abstract

Training practitioners do not know much about the large research literature on learning and instruction and are not current in the theory of cognition and learning that underlies the research. Yet they are responsible for determining what to train and how, and for deciding how to assess the progress of student learning. They develop, design, and implement training courses and curricula. They write the books and manuals. The purpose of this project is to provide research syntheses that recommend training practices that work effectively and to develop a better avenue for transitioning research knowledge into practice. These summaries will be published as a special issue of the *International Journal of Educational Technology*. This Journal provides a forum for communicating with a large audience of training practitioners and managers.

Background

The military services (and much of industry) rely on subject-matter specialists/experts for developing training. These individuals are seldom more than novices about the scientific research knowledge that provides the basis for choosing representations, methods of presentation, learning assessment procedures, and fostering the development of competency in adult learners (cf Latham, 1988). Yet they write the books and other training materials, design the tests, manage the development of competency, or approve the quality of contractor developed materials. A detailed knowledge and understanding of the research literature on cognition, learning, and instructional technology would enhance performance of these tasks. Therefore, any mechanism that could communicate this information effectively would contribute to improving the quality of training.

Training system practitioners, in part, try to overcome some of the deficits by hiring training or educational specialists to help. Although they do provide an important resource that improves the system, the adequacy of their training is questionable (there

are few focused programs to train them nationwide), and there are relatively few of them available. Thus, the bulk of the work preparing instructional materials, and decisions about their adequacy or quality, is done by individuals relatively uninformed about the empirical and theoretical basis for instructional techniques. How can we transition into practice the knowledge about different methods' effectiveness?

We adapted an approach to help solve this problem, which was first used by the Department of Education (Bennett, 1986). The approach provides summaries of information based on a synthesis of research in an easily understandable, accurate form about what works in educating and training students. The Department of Education document became the most widely distributed document on instructional research ever. It was targeted primarily at parents and teachers of young children attending schools in the U.S. The focus of our document is on instructing young adults attending military schools. The summaries provide a source of information to guide executives who manage and make policy, instructors and training specialists who are curriculum designers, developers, and evaluators. This information was presented as a document (Montague, 1988) published as a Navy Education and Training Command report and distributed widely among Navy training organizations.

Recently, because of the success of the approach, we were asked to refine it further. We were invited to prepare a special issue of the *International Journal of Educational Research* based on the original Navy document and to report on the reactions and use of the document in Navy training organizations. This project provided partial support for preparing this document.

Objectives

As a technological development, the purpose of the present effort is to explore the usefulness of research syntheses for transitioning research knowledge into practice, to develop a special issue displaying the research syntheses, and to discuss their utility.

Approach

Research syntheses were brought up to date and revised, based on comments on the earlier issue. Several new syntheses were prepared. Articles were solicited from professionals discussing the adequacy and utility of such syntheses for guiding instruction and instructional development. Data were obtained on the use of the materials in Navy training and are presented in an article.

Progress

A new document was drafted during the last half of FY91 entitled: *What works in adult instruction: The management, design, and delivery of instruction.* It is being edited and revised and its expected completion date is early 1992, with publication soon thereafter.

In one chapter, we list the organizations using the document and report the comments of training managers, specialists, and instructors regarding its use. An extensive discussion regarding the perceived 'obviousness' of generalizations for social/behavioral science research was included to confront comments that the information is common sense knowledge and to explain that perception is deceiving. We will record the use of the special issue and will include a response form for users to notify us about the material presented in the Special Issue.

References

Bennett, W. J. (1986). *What works: Research about teaching and learning.* Washington, DC: U.S. Department of Education.

Latham, G. P. (1988). Human resource training and development. *Annual Review of Psychology, 39*, 545-582.

Montague, W. E. (Ed.) (1988). *What works: Summary of research findings with implications for Navy instruction and learning* (NAVEDTRA 115-1). Pensacola, FL: Chief of Naval Education and Training.

An Examination of Cognitive and Motivational Effects of Employee Interventions

John P. Sheposh

Abstract

Total Quality Leadership (TQL) requires a fundamental change in the way work is conducted and an organization operates. The major focus of the present research is on the effect of TQL on worker empowerment and intrinsic motivation. A model was designed to test the following relationships: (1) Specific TQL related impediments and organizational climate influence the acceptance and use of TQL, (2) TQL acceptance and use has an impact on properties of individual jobs and employees' perception of their jobs, (3) these effects combine to influence job satisfaction and perceived stress. Data were collected on surveys in two Navy engineering facilities. The 368 subjects were engaged in jobs that were primarily scientific and technical in nature. Results indicated that the strongest impediments to successful implementation are associated with an organizational climate that is inconsistent with TQL principles. In addition, the results were generally consistent with the proposed model thus providing support for the contention that the employee's sense of empowerment and intrinsic motivation mediates the relationship between TQL and levels of satisfaction and stress.

Background

Public and private sector organizations are showing continuing interest in adopting Total Quality Management as a system to improve organizational performance. This is clearly evident from the Navy-wide implementation of TQL that is presently taking place. Total quality programs are characterized by the following: a focus on systemic rather than individual causes of poor quality, the use of statistical evidence as the basis for quality improvement actions and for the assessment of their impact, an emphasis on intra- and inter-departmental communication in solving and preventing problems, and removal of defects through process improvement rather than through inspection (Deming, 1982). Full-scale adoption of these principles involves a major change in the organization's orientation toward the way work is conducted. TQL is far from

a cosmetic fix. It requires a fundamental change in the way work is conducted and the way an organization operates.

Considering the nature of the changes dictated by a quality improvement intervention (e.g., required training, changes in work procedures) and the effort required of individual workers to enact these changes, it is extremely important to examine the way in which the quality program affects the individual employee. The emphasis in selling these programs, understandably, is on the advantages they give the organization. Solely restricting the focus to the benefits to the organization (e.g., profitability, higher productivity, and improved client satisfaction), however, may create some serious difficulties in selling quality improvement programs to employees (c.f. Guaspari, 1987). By only stressing organizational benefits, the employee may perceive that he/she is being asked to change, to work harder, to be more closely monitored, and to achieve goals of increasing difficulty without personal rewards. This may seriously compromise the implementation effort. It follows that research on the effects of quality programs on the individual would be beneficial to the assessment and implementation of such quality programs.

TQL, when properly implemented, should alter a person's job in significant ways (e.g., collecting and reporting data, focus on process, increased interaction with supervisors, co-workers and management, increased responsibility, and more precise and frequent job feedback). Ideally, these changes provide workers the opportunity to participate in decision-making and enhance personal reliance, a sense of autonomy and control which theoretically should lead to heightened intrinsic motivation and empowerment (Conger & Kanungo, 1988). The effect of TQL on worker empowerment and intrinsic motivation is the major focus of the present research.

In order to test the effect of TQL on workers' sense of empowerment and intrinsic motivation of a model, which incorporates elements of previous models (e.g., Hackman & Oldham, 1980, Tymon, 1988) as well as information concerning the organization and the implementation, is proposed (see Figure 1). The model is designed to test the following proposed relationships: (1) specific TQL-related impediments and organizational climate factors influence the degree to which TQL is successfully implemented and practiced, (2) the status of TQL in terms of implementation and use in turn is expected to affect properties of individual jobs and employees' perceptions

of their jobs, and (3) these effects will combine to influence the level of job satisfaction, perceived stress, and job performance. Because the proposed model examines the components of the Hackman and Oldham (1980) and Tymon (1988) models as well as organizational factors relevant to the intervention (TQL), it is considered to be more comprehensive and more capable of describing and explaining the effect of TQL on workers' jobs and resultant performance.
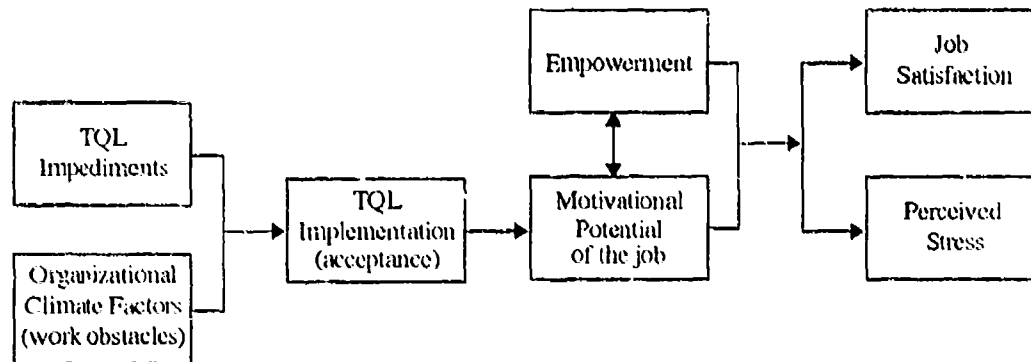
Figure 1. Proposed model.

A 3 year longitudinal study is being conducted to gain a better understanding of the effect TQL has on individual employees' jobs over time with special emphasis on intrinsic job rewards, job impediments, and personal control over work processes (empowerment), and to compare the proposed model with those of Hackman and Oldham (1980) and Tymon (1988) to determine which is the best explanatory model. The primary objectives of the work conducted in FY91 were to determine the acceptance of TQL and the organizational factors and impediments affecting acceptance and use, and to determine the correspondence between the data obtained for the first year and the proposed model.

Approach

A survey instrument is the primary means of data collection used to assess the effect of TQL on individual jobs, perceived stress, and job satisfaction. The areas assessed in the survey correspond to the major components of the proposed model: (1) general organizational climate characteristics (e.g., openness, communication, cooperation), (2) impediments affecting the

successful implementation and use of TQL (e.g., fear, lack of support, lack of adequate training), (3) personal and organizational acceptance of TQL, (4) specific job characteristics (e.g., skill variety, task significance), (5) empowerment (e.g., accomplish objectives), (6) job satisfaction, and (7) perceived job and organizational stress (e.g., lack of control over job decisions).

Two U.S. Navy engineering facilities, Naval Ship Weapon Systems Engineering Station (NSWSES) and the Integrated Combat Systems Engineering Station (ICSTF) are serving as the test sites. A first administration of the survey questionnaire has been conducted at both sites. Two hundred and ninety-five randomly selected employees from NSWSES and all 73 employees from ICSTF completed the survey.

Results

Overall the results from the first survey administration indicate that the strongest impediments to successful implementation were associated with an organizational climate inconsistent with TQL principles (e.g., poor fit between current policies and TQL). Results also indicate that the majority of the survey respondents were positively disposed toward TQL but reported that organizational support was less positive. Personal acceptance of TQL was most strongly associated with TQL involvement. Organizational acceptance of TQL was most strongly affected by levels of perceived impediments and organizational climate characteristics.

In addition, the data were generally consistent with the proposed model with some modification (see Figure 2). Organizational climate and TQL affected empowerment and the motivating properties of one's job, and these variables in turn affected the level of reported job satisfaction and stress. In contrast to the proposed model, empowerment is significantly related to job satisfaction but not to job stress while the motivating potential of the job is significantly linked solely to job stress. These preliminary results suggest that TQL has an impact on one's perceived sense of empowerment and intrinsic motivation of the job and that the proposed model successfully captures these interrelationships.

Continuation of this study will provide information about the effect of TQL over time on individual employees' jobs that

Figure 2. Model based on path analyses.

could not be accomplished from a 1 year study. To determine the incentive value of TQL over time, these data will be analyzed by means of a latent variable model with a minimum of two time periods (Williams & Podsakoff, 1989). From both a theoretical and applied perspective, a longitudinal test of the model components will provide a fuller understanding of the reciprocal causal relationships of TQL and employee attitudes, motivation, and empowerment. Additionally, information obtained from an extended period would enable DoD organizations to better understand the implementation issues involved in the adoption of TQL as well as the solutions to problems this approach may pose for their employees.

Publications and Presentations

Sheposh, J. P., Shettel-Neuber, J., Rosenthal, M., & Heller, C. (1991). *Impediments to TQL implementation within a U.S. Navy activity.* Paper presented at the annual conference of the Military Testing Association, San Antonio, TX.

Rosenthal, M., Sheposh, J. P., & Shettel-Neuber, J. (1992). *The role of organizational constraints, job stress, and efficacy on job satisfaction.* Paper to be submitted for review at the convention of the American Psychological Society, San Diego, CA.

References

Conger, J. A., & Kanungo, R. N. (1988). The empowerment process: Integrating theory and practice. *Academy of Management Review, 27*, 454-489.

Deming, W. E. (1982). *Quality, productivity, and competitive position.* Cambridge, MA: MIT, Center for Advanced Engineering Study.

Guaspari, J. (1987). The role of human resources in "selling" quality improvement to employees. *Management Review*, 20-24.

Hackman, J. R., & Oldham, G. R. (1980). *Work Redesign.* Reading, MA: Addison-Wesley.

Tymon, W. G., Jr. (1988). *An empirical investigation of a cognitive model of empowerment* (unpublished doctoral dissertation). Philadelphia, PA: Temple University.

Williams, L. J., & Podsakoff, P. M. (1989) Longitudinal field methods for studying reciprocal relationships in organizational behavior research: Toward improved causal analysis. In L.L. Cummings & B.M. Staw (Eds.), *Research in organizational behavior* (Vol II, pp. 247-292). Greenwich, CT: JAI Press.

Distribution List

Chief of Naval Personnel

Assistant for Manpower, Personnel, and Training (OP-987H)

Annapolis Technical Information Center, D. W. Taylor Naval Ship R&D Center

Commanding Officer, Naval Coastal Systems Center (Code 2120)

Commanding Officer, Naval Underwater Systems Center (Code 102)

Commander, Naval Command, Control and Ocean Surveillance Center, Research, Development, Test and Evaluation Division (Code 0141)

Commander, Naval Air Development Center (Code 01B)

Officer in Charge, Naval Surface Warfare Center Detachment, White Oak Laboratory (Code R04, Code H023)

Commander, Naval Weapons Center, Technical Library (Code 343)

Office of Naval Research (OCNR-10P6)

Office of Naval Research, Detachment London

Office of Naval Technology (OCNR-20P4)

Chief of Naval Education and Training (Code L01)

Commanding Officer, Naval Aerospace Medical Research Laboratory (Code 26)

Commander, U.S. ARI, Behavioral and Social Sciences (PERI-POT-I)

Department of the Air Force, DET5, Armstrong Laboratory Directorate, Brooks Air Force Base, TX

Program Manager, Life Sciences Directorate, Bolling Air Force Base (AFOSR/NRL)

Commander, Air Force Human Resources Laboratory, Wright-Patterson Air Force Base (AFHRL/LRS-TDC)

Department of the Air Force (AFSC), Air Force Human Resources Laboratory, Lowry Air Force Base (AFHRL/ID)

Superintendent, Naval Postgraduate School

Director of Research, U.S. Naval Academy

Institute for Defense Analyses, Science and Technology Division

Defense Technical Information Center (DTIC)

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE March 1992	3. REPORT TYPE AND DATE COVERED Administrative Publication--Oct 90-91

4. TITLE AND SUBTITLE
Independent Research and Independent Exploratory Development
FY91 Annual Report

5. FUNDING NUMBERS
Center Overhead

6. AUTHOR(S)
William E. Montague (Editor)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Navy Personnel Research and Development Center
San Diego, California 92152-6800

8. PERFORMING ORGANIZATION REPORT NUMBER
NPRDC-AP-92-5

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Technology/Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT (Maximum 200 words)

This report documents 6.1 and 6.2 research efforts conducted at the Navy Personnel Research and Development Center under the Independent Research/Independent Exploratory Development (IR/IED) program. The FY91 IR program includes the following project articles: The Role of Feedback in Computer-based Training; Exploratory Examination of Artificial Neural Networks as an Alternative to Linear Regression; Artificial Neural Networks and Training; A Comparison of Artificial Neural Networks and Linear Regression; Using Neural Networks to Predict Behavior from Psychological Data; and Experimentally-based Learning of Multiple Roles. Progress reports on IED projects include: A Large-scale Model to Maximize Navy Unit Personnel Readiness; Effects of Administration Method and Anonymity/Identification on Survey Responses; An Examination of Cognitive and Motivational Effects of Employee Interventions; and Transitioning Research Knowledge into Practice.

14. SUBJECT TERMS
Independent Research, Independent Exploratory Development, IR/IED

15. NUMBER OF PAGES
94

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UNLIMITED